(19) BUNDESREPUBLIK DEUTSCHLAND

DEUTSCHES PATENTAMT

(12) **Patentschrift**

(10) **DE 195 33 541 C 1**

(51) Int. Cl.[6]:
G 10 L 7/08
B 60 R 11/02
B 60 R 16/02
H 02 J 13/00
G 01 C 21/00
G 08 G 1/0968
H 03 J 1/00
G 09 B 29/10

(21) Aktenzeichen: 195 33 541.4-53
(22) Anmeldetag: 11. 9. 95
(43) Offenlegungstag: —
(45) Veröffentlichungstag der Patenterteilung: 27. 3. 97

Innerhalb von 3 Monaten nach Veröffentlichung der Erteilung kann Einspruch erhoben werden

(73) Patentinhaber:

Daimler-Benz Aerospace Aktiengesellschaft, 80804 München, DE; Daimler-Benz Aktiengesellschaft, 70567 Stuttgart, DE; Mercedes-Benz Aktiengesellschaft, 70327 Stuttgart, DE

(72) Erfinder:

Stammler, Walter, Dr., 89077 Ulm, DE; Nüßle, Gerhard, Dipl.-Ing., 89134 Blaustein, DE; Class, Fritz, Dr., 72587 Römerstein, DE; Möller, Karsten-Uwe, 73730 Esslingen, DE; Reh, Frank, 70439 Stuttgart, DE; Buschkühl, Burkard, 71069 Sindelfingen, DE; Heinrich, Christian, Dr., 72733 Esslingen, DE

(56) Für die Beurteilung der Patentfähigkeit in Betracht gezogene Druckschriften:

DE 39 28 049 A1
DE 38 19 178 A1
CLASS, F., KATTERFELDT, P., REGEL, P.: »Methoden und Algorithmen der Worterkennung«. In: MANGOLD, H. (Herausgeber) Sprachliche Mensch-Maschine-Kommunikation, Verlag Oldenbourg 1992, S. 1-13;
SHINOHARA, T., MAEDA, N., ASADA, H.: »Hands Free Voice Recognition Telephone For Automobile«. In: Proceedings of the ISATA-Conference 1990, S. 525-545;
ASADA, H., NORIMATSU, H., AZUMA, S.: »Speaker-Dependent Voice Recognition Algorithm For Voice Dialing In Automotive Environment«. In: Proceedings of the ISATA-Conference 1990, S. 547-557;

(54) Verfahren zur automatischen Steuerung eines oder mehrerer Geräte durch Sprachkommandos oder per Sprachdialog im Echtzeitbetrieb und Vorrichtung zum Ausführen des Verfahrens

(57) Die Erfindung betrifft ein Sprachbediensystem, bei dem ein Verfahren zur automatischen Steuerung von Geräten per Sprachdialog angewendet wird, das auf Verfahren zur Sprachausgabe, Sprachsignalvorverarbeitung und Spracherkennung, syntaktisch-grammatikalischer Nachverarbeitung sowie Dialog-, Ablauf- und Schnittstellensteuerung basiert und dadurch gekennzeichnet ist, daß
- Syntax- und Kommandostruktur während des Echtzeit-Dialogbetriebs fixiert sind,
- Vorverarbeitung, Erkennung und Dialogsteuerung für Betrieb in geräuschbehafteter Umgebung ausgelegt sind,
- für die Erkennung allgemeiner Kommandos kein Training durch den Benutzer erforderlich ist,
- für die Erkennung spezifischer Kommandos einzelner Benutzer ein Training notwendig ist,
- die Eingabe von Kommandos verbunden erfolgt, wobei die Anzahl der Worte, aus denen ein Kommando für die Spracheingabe gebildet wird, variabel ist,
- eine echtzeitige Verarbeitung und Abwicklung des Sprachdialoges gegeben ist,
- die Sprachein- und -ausgabe im Freisprechbetrieb erfolgt.

DE 195 33 541 C1

Beschreibung

Die Erfindung betrifft ein Verfahren zur automatischen Steuerung eines oder mehrerer Geräte durch Sprachkommandos oder per Sprachdialog im Echtzeitbetrieb gemäß Oberbegriff des Patentanspruchs 1 sowie eine Vorrichtung zum Ausführen des Verfahrens gemäß Oberbegriff des Patentanspruchs 50.

Ein solches Verfahren ist bereits aus der DE 38 19 178 A1 bekannt.

Verfahren bzw. Vorrichtungen dieser Art werden allgemein in sogenannten Sprachdialog- bzw. Sprachbediensystemen z. B. für Fahrzeuge, Computer, Roboter, Maschinen, Anlagen usw. eingesetzt.

Ein Sprachdialogsystem (SDS) läßt sich i.a. im wesentlichen auf folgende Komponenten reduzieren (vgl. hierzu z. B.: F. Class, H. Katterfeldt, P. Regel: "Methoden und Algorithmen der Worterkennung"; in: H. Mangold (Herausgeber): Sprachliche Mensch-Maschine-Kommunikation (Verlag Oldenbourg, 1992), Seiten 1 bis 13):

- Spracherkennungssystem, welches ein eingesprochenes Kommando ("Sprachkommando") mit anderen erlaubten Sprachkommandos vergleicht und eine Entscheidung trifft, welches Kommando aller Wahrscheinlichkeit nach eingesprochen wurde,
- Sprachausgabe, welche die zur Benutzerführung erforderlichen Sprachbefehle und Signalisierungstöne ausgibt und ggf. das Erkennergebnis rückmeldet,
- Dialog- und Ablaufsteuerung, um dem Benutzer zu verdeutlichen, welche Art von Eingabe erwartet wird, bzw. um zu prüfen, ob die erfolgte Eingabe konsistent ist mit der Aufforderung und mit dem momentanen Status der Applikation, und um die resultierende Aktion bei der Applikation (z. B. dem zu steuernden Gerät) anzustoßen,
- Kontrollinterface als Schnittstelle zur Applikation: Dahinter verbergen sich Hard- und Softwaremodule, um verschiedene Aktuatoren bzw. Rechner anzusteuern, die die Applikation beinhalten,
- Applikation, die per Sprache angesteuert wird: Das kann z. B. ein Bestell- oder Auskunftsystem, ein CAE-Arbeitsplatz oder ein behindertengerechter Rollstuhl sein.

Die vorliegende Beschreibung konzentriert sich — ohne Einschränkung auf die allgemeine Anwendbarkeit der geschilderten Verfahren, Vorrichtungen und Abläufe — auf die Spracherkennung, die Dialogstruktur sowie auf eine spezielle Applikation in Kraftfahrzeugen.

Die Schwierigkeiten bei bisher bekannten Lösungen (vgl.: z. B. T. Shinohara, N. Maeda, H. Asada: "Hands Free Voice Recognition Telephone For Automobile"; in: Proceedings of the ISATA-Conference 1990, Seiten 525 bis 545 sowie H. Asada, H. Norimatsu, S. Azuma: "Speaker-Dependent Voice Recognition Algorithm For Voice Dialing In Automotive Environment"; in: Proceedings of the ISATA-Conference 1990, Seiten 547 bis 557) liegen

a) in der Notwendigkeit, ein aufwendiges Training durchzuführen, um das System auf die Charakteristik des jeweiligen Sprechers oder auf einen wechselnden Wortschatz anzupassen. Die Systeme sind in der Regel entweder vollständig sprecherunabhängig oder vollständig sprecherabhängig bzw. sprecheradaptiv, wobei letztere für jeden neuen Benutzer einen Trainingslauf erfordern. Dies kostet Zeit und reduziert den Bedienkomfort bei häufig wechselnden Sprechern sehr stark. Aus diesem Grund ist bei herkömmlichen Systemen auch der Vokabularumfang gering bei Applikationen, wo mit wechselnden Sprechern und Zeitnot der einzelnen Sprecher zu rechnen ist,

b) in dem unzureichenden Bedienkomfort, der darin zum Ausdruck kommt, daß
   - das Vokabular auf ein Minimum begrenzt ist, um hohe Erkennsicherheit zu garantieren,
   - die Einzelworte eines Kommandos isoliert (d. h. mit Zwischenpausen) eingegeben werden,
   - Einzelworte quittiert werden müssen, um Fehler zu erkennen,
   - mehrstufige Dialoghierarchien abzuarbeiten sind, um vielfältige Funktionen zu steuern,
   - ein Mikrofon in die Hand zu nehmen ist bzw. ein Headset getragen werden muß,

c) in der fehlenden Robustheit
   - gegenüber Bedienfehlern
   - gegenüber störenden Umgebungsgeräuschen,

d) in der aufwendigen und teueren Hardware-Realisierung, vor allem bei mittleren und kleinen Stückzahlen.

In der eingangs bereits genannten DE 38 19 178 A1 wird ein Spracherkennungssystem beschrieben, bei dem die eingegebenen Sprachkommandos mittels eines sprecherunabhängigen Verbundwort-Spracherkenners und eines sprecherabhängigen Zusatz-Spracherkenners erkannt und gemäß ihrer Erkennungswahrscheinlichkeit klassifiziert werden.

Bei diesem Spracherkennungsverfahren wird zuerst ein unbekanntes Sprachkommandomuster aus Merkmalen erzeugt, welche aus dem unbekannten Sprachkommando extrahiert worden sind. Danach wird ein Ähnlichkeitsgrad zwischen dem erzeugten unbekannten Muster und Referenzmustern ermittelt, die sich zusammensetzen

a) aus Referenzmustern, die ausschließlich für eine sprecherunabhängige Erkennung verwendet worden sind, und

b) aus Referenzmustern, die ausschließlich für eine sprecherabhängige Erkennung verwendet worden sind.

Anschließend wird der Ähnlichkeitsgrad jedes Referenzmusters entweder bezüglich der sprecherunabhängigen oder bezüglich der sprecherabhängigen Erkennung korrigiert, indem der ermittelte Ähnlichkeitsgrad einer vorgegebenen Operation unterzogen wird. Danach wird das Muster mit dem höchsten Ähnlichkeitsgrad bestimmt.

Die zugehörige Spracherkennungseinrichtung weist eine Koeffizientenspeichereinrichtung auf, um den erhaltenen Ähnlichkeitsgrad entsprechend zu korrigieren, sowie eine Sprachidentifizierungseinrichtung, um die Ähnlichkeitsgrade des Musters, das entweder bei einer sprecherunabhängigen oder bei einer sprecherabhängigen Erkennung geliefert worden ist, mit korrigierten Ähnlichkeitsgraden des Musters zu vergleichen und um das Muster mit dem höchsten Ähnlichkeitsgrad zu bestimmen. Das System kann per Sprachkommando oder per Sprachdialog betrieben werden.

Aus der DE 39 28 049 A1 ist ein Verfahren zur automatischen Steuerung eines Archivierungssystems durch Sprachkommandos bekannt, bei dem erkannte zulässige Sprachkommandos auf ihre Plausibilität hin überprüft

werden.

Die Aufgabe der Erfindung besteht darin, zum einen ein Verfahren anzugeben, mit dem mit möglichst geringem Aufwand ein oder mehrere Geräte durch Sprachkommandos oder per Sprachdialog zuverlässig im Echtzeitbetrieb gesteuert werden können. Ferner soll eine geeignete Vorrichtung angegeben werden, mit der das zu schaffende Verfahren ausgeführt werden kann.

Die erfindungsgemäße Lösung der Aufgabe ist in bezug auf das zu schaffende Verfahren durch die Merkmale des Patentanspruchs 1 und in bezug auf die zu schaffende Vorrichtung durch die Merkmale des Patentanspruchs 50 wiedergegeben. Die übrigen Ansprüche enthalten vorteilhafte Aus- und Weiterbildungen des erfindungsgemäßen Verfahrens (Ansprüche 2 bis 49) sowie der erfindungsgemäßen Vorrichtung (Ansprüche 51 bis 62).

Der wesentliche Vorteil der Erfindung ist darin zu sehen, daß mit relativ geringem Aufwand eine zuverlässige Steuerung bzw. Bedienung von Geräten per Sprachkommando bzw. per Sprachdialog im Echtzeitbetrieb möglich ist.

Ein weiterer wesentlicher Vorteil ist darin zu sehen, daß eine der natürlichen Sprechweise weitgehend angepaßte Eingabe der Sprachkommandos bzw. Führung des Sprachdialogs mit dem System möglich ist und daß dem Sprecher hierfür ein umfangreiches Vokabular von zulässigen Kommandos zur Verfügung steht.

Ein dritter Vorteil ist darin zu sehen, daß das System fehlertolerant arbeitet und in einer vorteilhaften Weiterbildung der Erfindung z. B. auch nichtzulässige Wörter, Namen, Laute oder Wortumstellungen in den vom Sprecher eingegebenen Sprachkommandos i.a. als solche erkennt und aus diesen eingegebenen Sprachkommandos von dem Sprecher an sich gewollte zulässige Sprachkommandos extrahiert.

Im folgenden wird die Erfindung anhand der Figuren näher erläutert. Es zeigt

Fig. 1 das Blockschaltbild einer bevorzugten Ausführungsform der erfindungsgemäßen Vorrichtung zum Ausführen des erfindungsgemäßen Verfahrens ("Sprachdialogsystem"),

Fig. 2 eine detaillierte Darstellung des eigentlichen Sprachdialogsystems gemäß **Fig. 1**,

Fig. 3 das Flußdiagramm zu einer bevorzugten Ausführungsform der Segmentierung der eingegebenen Sprachkommandos für ein Sprachdialogsystem gemäß **Fig. 2**,

Fig. 4 und 5 Ausführungsbeispiele von Hidden-Markov-Modellen,

Fig. 6 den hardwaremäßigen Aufbau eines bevorzugten Ausführungsbeispiels des Sprachdialogsystems gemäß **Fig. 2**,

Fig. 7 das Zustandsdiagramm für die Anwendung des Sprachdialogsystems gemäß **Fig. 2** zur sprachgesteuerten Bedienung eines Telefons,

Fig. 8 das Flußdiagramm zur Bedienung eines Telefons gemäß **Fig. 7**,

Fig. 9 und 10 das Flußdiagramm zur Funktion "Namenswahl" (**Fig. 9**) bzw. "Nummernwahl" (**Fig. 10**) bei der Bedienung eines Telefons gemäß Flußdiagramm nach **Fig. 8**.

Das im folgenden beschriebene Sprachdialogsystem (SDS) in **Fig. 1** umfaßt die Komponenten Spracheingabe (symbolisch dargestellt durch ein Mikrofon), Spracherkennung, Dialog- und Ablaufsteuerung, Kommunikations- und Kontrollinterface Sprachausgabe mit angeschlossenem Lautsprecher sowie (beispielhaft) eine Applikation, d. h. ein durch das SDS zu steuerndes bzw. zu bedienendes Gerät. SDS und Applikation bilden zusammen ein Sprachbediensystem (SBS), das in Echtzeit ("online") betrieben wird.

Die Syntax- und Dialogstruktur und die für alle Benutzers Sprecher verbindlichen Basissprachkommandos werden "offline" außerhalb des SDS bzw. SBS (beispielhaft) mit Hilfe einer PC-Workstation im "off-line Dialog Editormodus" erstellt und fixiert und zusammen mit vorzugebenden Parametern und Ablaufstrukturen dem SDS bzw. SBS vor Inbetriebname in Form von Datenfiles übergeben.

Das SDS der **Fig. 1** ist in **Fig. 2** im Detail dargestellt. Ein (nicht gezeigtes) Mikrofon ist mit einem Analog/Digital-Wandler verbunden, der über Vorrichtungen zur Geräuschreduktion, Echokompensation und Segmentierung mit einem sprecherunabhängigen Verbundwort-Spracherkenner und mit einem sprecherabhängigen Spracherkenner verbunden ist. Die beiden Spracherkenner sind ausgangsseitig mit einer Einheit zur syntaktisch-grammatikalischen und semantischen Verarbeitung der Erkenner-Ausgangssignale verbunden. Diese Einheit wiederum ist mit der Dialog- und Ablaufsteuerung verbunden, die ihrerseits zum einen über Schnittstellen (z. B. D2B, V24, CAN, PCMCIA usw.) mit den (nicht gezeigten) Geräten verbunden ist, die über das SDS angesteuert bzw. bedient werden sollen. Die Dialog- und Ablaufsteuerung ist ferner mit einer Spracheingabe-/Sprachausgabe-Einheit verbunden, die aus einem Sprachencoder, einem Sprachdecoder und einem Sprachspeicher besteht.

Der Sprachencoder ist eingangsseitig an den Ausgang der Vorrichtung zur Geräuschreduktion und ausgangsseitig an den Sprachspeicher angeschlossen. Der Sprachspeicher ist ausgangsseitig an den Sprachdecoder angeschlossen, der ausgangsseitig über einen Digital/Analog-Wandler mit einem (nicht gezeigten) Lautsprecher verbunden ist.

Die Vorrichtung zur Echokompensation ist über Schnittstellen mit (nicht gezeigten) Geräten/Sensoren verbunden, die ggf. zu kompensierende Audiosignale liefern.

Der sprecherunabhängige Verbundwort-Spracherkenner weist zum einen eine Einheit zur Merkmalsextraktion auf, in der die Cepstrumsbildung und die Adaption des Erkenners u. a. an die analoge Übertragungscharakteristik der eingehenden Signale durchgeführt werden, und zum anderen eine nachgeschaltete Einheit zur Klassifikation.

Der sprecherabhängige Spracherkenner weist ebenfalls zum einen eine Einheit zur Merkmalsextraktion und zum anderen ein Einheit zur Klassifikation auf. Zusätzlich ist jedoch über einen Umschalter anstelle der Klassifikationseinheit eine Einheit zur Eingabe der sprecherspezifischen Zusatzsprachkommandos zuschaltbar, die in den Trainingsphasen vor, während oder nach dem Echtzeitbetrieb des SDS vom Erkenner trainiert werden sollen. Der sprecherabhängige Erkenner arbeitet z. B. nach dem Dynamic-Time-Warping(DTW)-Verfahren, nach dem dessen Klassifikationseinheit die Abstände zwischen dem zu erkennenden Kommando und vortrainierten Referenzmustern feststellt und das Referenzmuster mit dem geringsten Abstand als das zu erkennende Kommando identifiziert. Alternativ hierzu kann aber auch der sprecherabhängige Erkenner mit Methoden der Merkmalsextraktion arbeiten, wie sie in sprecherunabhängigen Spracherkennern zur Anwendung kommen (Cepstrumsbildung, Adaption usw.).

Im folgenden wird die Funktionsweise des SDS näher erläutert.

Das SDS beinhaltet — wie zuvor ausgeführt — zweierlei Spracherkennertypen zur Erkennung vorgegebener Sprachkommandos. Die beiden Erkenner können wie folgt charakterisiert werden:

- Sprecherunabhängige Erkennung von verbunden gesprochenen Worten. Damit lassen sich allgemeine Steuerkommandos, Ziffern, Namen, Buchstaben etc. erkennen, ohne daß der Sprecher bzw. Benutzer eines oder mehrere der benutzten Worte vorher trainiert haben muß.

Weiterhin kann die Eingabe im Verbundwortmodus erfolgen, d. h. eine Kombination mehrerer Worte, Ziffern, Namen ergibt ein Kommando, welches in einem Zug, d. h. ohne Pause eingesprochen wird (z. B. das Kommando: "Kreis mit Radius Eins"). Beim Algorithmus zur Klassifikation handelt es sich um einen HMM(Hidden-Markov-Modell)-Erkenner, der im wesentlichen auf Phonemen (Lautuntereinheiten) aufbaut und daraus Worte bzw. Kommandos zusammensetzt. Das Vokabular und die daraus aufgebauten Kommandos ("Syntaxstruktur") werden vorab im Labor fixiert und dem Erkenner in Form von Datenfiles übergeben ("off-line Dialog Editiermodus"). Im Echtzeit-Betrieb kann das Vokabular und die Syntaxstruktur des unabhängigen Erkenners vom Benutzer nicht modifiziert werden.

- Sprecherabhängige Erkennung von benutzer-/sprecherspezifischen Namen oder Funktionen, die der Benutzer/Sprecher definiert und trainiert.

Der Benutzer/Sprecher hat die Möglichkeit, ein persönliches Vokabular in Form von Namenslisten, Funktionslisten etc. anzulegen bzw. zu editieren. Dadurch kann der Benutzer/Sprecher seinen persönlichen Wortschatz wählen und diesen jederzeit "on line" d. h. im Echtzeitbetrieb, an seine Bedürfnisse anpassen.

Als Beispiel für eine Anwendung im Telefonumfeld sei die "Namensliste" genannt, d. h. das individuelle Verzeichnis von Namen, wobei

— der Namen in einer Trainingsphase ein- oder mehrmals vom Benutzer eingesprochen wird (z. B. "Onkel Willi") und dem Namen per Tastatureingabe, vorzugsweise aber per unabhängigem Spracherkenner eine Telefonnummer zugeordnet wird,
— nach Abschluß des obigen Trainings und der Nummernzuweisung der Benutzer nur noch dem sprecherabhängigen Erkenner einen Namen ("Onkel Willi") nennt, nicht aber die zugehörige Telefonnummer, die dem System bereits bekannt ist.

Der sprecherabhängige Erkenner wird in der

— einfachsten Form als Einzelworterkenner ausgelegt
— in der leistungsfähigeren Form als Verbundworterkenner, der nahtlos mit dem sprecherunabhängigen Erkenner gekoppelt ist. ("Onkel Willi anrufen" als vollständiges Kommando, wobei das Wort "anrufen" Teil des sprecherunabhängigen Vokabulars ist).

Im Anschluß an die Spracherkennung wird eine Nachverarbeitung der mit einer bestimmten Erkennwahrscheinlichkeit behafteten Ergebnisse der beiden Spracherkenner durchgeführt.

Der sprecherunabhängige Verbundwort-Spracherkenner z. B. liefert mehrere Satzhypothesen in einer Reihenfolge, welche die Erkennwahrscheinlichkeiten repräsentiert. Diese Satzhypothesen berücksichtigen bereits die erlaubte Syntaxstruktur, d. h. innerhalb der syntaktischen Nachverarbeitung (Fig. 2) werden unzulässige Wortfolgen ausgesondert bzw. nach verschiedenen Kriterien bewertet, wie wahrscheinlich die hierin auftretende Wortkombination ist. Ferner werden die von den Spracherkennern erzeugten Satzhypothesen auf ihre semantische Plausibilität überprüft und danach die Hypothese mit der höchsten Wahrscheinlichkeit ausgewählt.

Ein korrekt erkanntes Sprachkommando wird an die Dialogsteuerung weitergereicht und führt anschließend zu einem diesem Sprachkommando zugeordneten Eingriff auf die Applikation, wobei die Meldung über das Kontrollinterface weitergereicht wird.

Das hier skizzierte System ist im "on-line"-Betrieb durch eine fixierte Syntax- und Kommandostruktur, sowie durch eine Kombination von fixiertem Vokabular (sprecherunabhängiger Erkenner) und frei definierbarem Vokabular, wie z. B. Namen (sprecherabhängiger Erkenner), gekennzeichnet.

Dieser zunächst starr erscheinende Rahmen ist eine Voraussetzung für hohe Erkennleistung bei großem Vokabularumfang (bis zu mehreren hundert Worten), bei geräuschbehafteter Umgebung, bei wechselnden akustischen Verhältnissen in der Fahrgastzelle sowie bei variierenden Sprechern. Der hohe Vokabularumfang wird genutzt, um die Benutzerfreundlichkeit durch Verwendung synonymer Worte bzw. unterschiedlicher Aussprachevarianten zu erhöhen. Auch erlaubt die Syntax die Umstellung von Worten in den Sprachkommandos, so z. B.:

"Größerer Radius bei linkem Kreis"
oder — alternativ hierzu —
"Bei linkem Kreis größerer Radius",

wobei diese Alternativen jedoch von vorneherein bei der Festlegung mit dem "off-line Dialog Editor" definiert sein müssen.

Der hier skizzierte Lösungsansatz erweist sich insofern als vorteilhaft, als

- die Verbundworteingabe von Kommandos natürlicher und schneller ist als die Isoliertworteingabe. Die Praxis hat insbesondere gezeigt, daß der unbefangene Benutzer schwer daran zu gewöhnen ist, abgehackt (mit deutlichen Zwischenpausen) zu sprechen, um ein Mehrwortkommando einzugeben (die Akzeptanz derartiger Systeme ist deshalb deutlich geringer),
- die Eingabe z. B. von Ziffern- oder Buchstabenkolonnen im Verbund leichter ist und weniger Konzentration erfordert als die Einzeleingabe,
- die Dialogführung natürlicher ist, weil z. B. bei Ziffernkolonnen nicht jede Einzelziffer quittiert werden muß, sondern nur der eingegebene Ziffernblock,
- wegen des Wortschatzes von z. B. bis zu einigen hundert Worten eine Vielzahl von Funktionen per Sprache bedienbar sind, die vorher manuelle Bedienung erforderten,
- die Menge manueller Schaltelemente reduziert werden kann bzw. bei Spracheingabe die Hände anderweitig benutzbar sind, z. B. bei der Qualitätskontrolle von Motoren.

Der Bedienkomfort wird bei dem vorliegenden System weiterhin erhöht durch Nutzung von Freisprechmikrofon(en) und Verzicht auf Headset (Kopfhörer und Lippenmikrofon) bzw. Handmikrofon. Das erfordert allerdings eine leistungsfähige Geräuschreduktion (Fig. 2) sowie ggf. eine Echokompensation von Signalen, die

z. B. aus dem Dialog- oder anderen Lautsprechern stammen.

Die vorgeschlagene Echokompensation erlaubt es insbesondere, der Sprachausgabe ins Wort zu fallen, d. h. den Erkenner anzusprechen, während die Sprachausgabe aktiv ist.

Gleichzeitig können im Labor per "off-line Dialog Editor" jederzeit das Vokabular und die Kommandos verändert werden, ohne daß dies ein neues Training mit einer Vielzahl von Sprechern für die neuen Worte des sprecherunabhängigen Erkenners bedarf. Der Grund liegt darin, daß im Labor die Datenbank für sprecherunabhängige Phoneme vorliegt und aus diesen Phonemen mit der vorhandenen Entwicklungsumgebung ohne weiteres neue Worte und Kommandos generiert werden können. Letztlich läuft eine Kommando- oder Vokabularänderung darauf hinaus, die im Labor mit dem Entwicklungssystem berechneten neuen Parameter und Daten als Datenfile in den sprecherunabhängigen "Echtzeit-Erkenner" zu überspielen und dort im Speicher abzulegen.

Mittels des vorgeschlagenen SDS können sowohl Funktionen innerhalb des Computers, in dem das SDS eingebaut ist, als auch externe Geräte bedient werden. Das SDS weist neben einer PCMCIA-Schnittstelle noch Schnittstellen auf, welche für externe Geräte zugänglich sind. Dies sind zum Beispiel:
V.24-Schnittstelle, optischer Daten-Steuerbus, CAN-Interface usw. Optional kann das SDS mit weiteren Schnittstellen ausgestattet werden.

Das SDS wird vorzugsweise durch Betätigen einer push-totalk-Taste (PTT) oder durch ein definiertes Schlüsselwort aktiviert. Die Abschaltung nach Beendigung eines Kommandos erfolgt automatisch durch die interne Segmentierung des SDS. In geräuscharmer Umgebung kann das SDS auch kontinuierlich aktiviert sein.

Ablaufbeschreibung

An dieser Stelle sei betont, daß das SDS in Fig. 2 nur ein Beispiel ist für ein nach der Erfindung mögliches SDS. Die Konfiguration der Schnittstellen zur Dateneingabe bzw. Datenausgabe bzw. zur Steuerung der angeschlossenen Komponenten ist hier ebenfalls nur beispielhaft dargestellt.

Die dargestellten Funktionsblöcke werden im folgenden näher erläutert.

1. Geräuschreduktion

Diese ermöglicht es, stationäre oder quasi-stationäre Umgebungsgeräusche vom digitalisierten Sprachsignal zu unterscheiden und diese vom Sprachsignal abzuziehen. Geräusche dieser Art sind z. B.: Fahrgeräusche in einem Kraftfahrzeug (Kfz), Umgebungsgeräusche in Labors und Büros wie Lüfter oder Maschinengeräusche in Fabrikationshallen.

2. Echokompensation

Über die Echokompensation werden die digitalisierten Lautsprechersignale z. B. der Sprachausgabe bzw. eines eingeschalteten Radios über adaptive Filteralgorithmen vom Mikrofonsignal subtrahiert. Die Filteralgorithmen bilden den Echopfad vom Lautsprecher zum Mikrofon nach.

3. Segmentierung

Die Segmentierung setzt — wie in Fig. 3 gezeigt — auf spektraltransformierten Daten auf. Hierzu werden die Signale blockweise zusammengefaßt (frame) und mit einer schnellen Fouriertransformation (FFT) in den Frequenzbereich umgesetzt. Durch Betragsbildung und Gewichtung mit einem gehörbezogenen MEL-Filter, d. h. einem dem melodischen Empfinden der Tonhöhe nachgebildeten Filter, bei dem eine gehörbezogene Einteilung des Sprachbereiches (~200 Hz bis ~6 kHz) in einzelne Frequenzbereiche ("Kanäle") durchgeführt wird, werden die Spektralwerte zu Kanalvektoren zusammengefaßt, die die Leistung in den verschiedenen Frequenzbändern angeben. Im Anschluß erfolgen eine Grobsegmentierung, die permanent aktiv ist und Kommandoanfang sowie Kommandoende grob erfaßt, sowie eine Feinsegmentierung, die im Anschluß daran die genauen Grenzen festlegt.

4. Merkmalsextraktion

Der Merkmalsextraktor berechnet aus den digitalisierten und segmentierten Sprachsignalen über mehrere Stufen hinweg Merkmalsvektoren und bestimmt den dazugehörigen normierten Energiewert.

Dazu werden beim sprecherunabhängigen Erkenner die Kanalvektoren mit einer diskreten Cosinustransformation (DCT) in Cepstralvektoren transformiert. Zusätzlich wird die Energie des Signals berechnet und normiert. Parallel dazu wird eine laufende Mittelwertberechnung der Cepstralwerte durchgeführt mit dem Ziel, den Erkenner sowohl an den momentanen Sprecher als auch auf Übertragungscharakteristiken, z. B. des Mikrofons und des Kanals (Sprecher → Mikrofon) zu adaptieren. Die Cepstralvektoren werden von diesem adaptierten Mittelwert befreit und mit der zuvor berechneten normierten Energie zu sogenannten CMF-Vektoren (Cepstralkoeffizienten mittelwertfrei) zusammengefaßt.

5. Klassifikation des sprecherunabhängigen Verbundwort-Spracherkenners

5.1 Hidden-Markov-Modell (HMM)

Ein Hidden-Markov-Modell ist eine Ansammlung von Zuständen, die untereinander durch Übergänge verbunden sind (Fig. 4).

Jeder Übergang, von einem Zustand q_i zum anderen q_j, wird durch eine sogenannte Übergangswahrscheinlichkeit beschrieben. Jedem Knoten (Zustand) ist ein Vektor von sogenannten Emissionswahrscheinlichkeiten der Länge M zugeordnet. Über diese Emissionswahrscheinlichkeiten wird die Verbindung zur physikalischen Welt hergestellt. Die Modellvorstellung geht soweit, daß in einem bestimmten Zustand q_i eines von M verschiedenen Symbolen "emittiert" wird, entsprechend der zustandsbezogenen Emissionswahrscheinlichkeit. Die Symbole stehen stellvertretend für die Merkmalsvektoren. Die Folge von "emittierten" Symbolen, die das Modell erzeugt, sind sichtbar. Die konkrete Abfolge der Zustände, die innerhalb des Modells durchlaufen werden, ist dagegen nicht sichtbar (engl. "hidden").

Ein Hidden-Markov-Modell ist durch folgende Größen definiert:

- T Anzahl der Symbole
- t Zeitpunkt für ein beobachtetes Symbol, t = 1 ... T
- N Anzahl der Zustände (Knoten) des Modells

- M Anzahl der möglichen Symbole (= Codebuchgröße)
- Q Zustände des Modells {q1, q2, ... qn}
- V Menge der möglichen Symbole
- A Übergangswahrscheinlichkeit vom einem Zustand in einen anderen
- B Wahrscheinlichkeit für ein Ausgabesymbol in einem Zustand des Modells (Emissionswahrscheinlichkeit)
- π Wahrscheinlichkeit für den Anfangszustand des Modells (beim Training des HMM's).

Unter Benutzung der Wahrscheinlichkeitsverteilungen A und B können mit Hilfe dieses Modells Ausgabesymbole erzeugt werden.

5.2 Aufbau des phonembasierten HMM-Erkenners

Bei einem Spracherkennungssystem mit größerem Wortschatz basiert die Worterkennung zweckmäßigerweise nicht auf Ganzwörtern, sondern auf phonetischen Wortuntereinheiten. Eine solche Wortuntereinheit ist zum Beispiel ein Laut, ein Diphon (Doppellaut) oder ein Lautübergang. Ein zu erkennendes Wort wird dann durch die Verkettung der entsprechenden Wortuntereinheiten-Modelle dargestellt. In Fig. 5 ist als Beispiel einer solchen Darstellung mit verketteten Hidden-Markov-Modellen zum einen die standardphonetische Beschreibung des Wortes "braten" (Fig. 5a) sowie zum anderen die phonetische Beschreibung von Aussprachevarianten (Fig. 5b) dargestellt. Diese Wortuntereinheiten-Modelle werden bei der Erstellung des Systems an Stichproben vieler Sprecher trainiert und bilden die Datenbasis, auf der der "offline Dialog Editor" auf setzt. Dieses Konzept mit Wortuntereinheiten hat den Vorteil, daß neue Wörter relativ einfach in das vorhandene Lexikon aufgenommen werden können, da die Parameter für die Wortuntereinheiten schon bekannt sind.

Theoretisch kann mit diesem Erkenner ein beliebig großes Vokabular e kannt werde. In der Praxis wird man jedoch durch beschränkte Rechenleistung und für die jeweilige Anwendung notwendige Erkennungsleistung an Grenzen stoßen.

Die Klassifikation basiert auf dem sogenannten Viterbialgorithmus, in welchem die Wahrscheinlichkeit jedes Wortes für die einlaufende Symbolfolge berechnet wird, wobei ein Wort hier als Verkettung verschiedener Phoneme zu verstehen ist. Der Viterbialgorithmus wird ergänzt durch eine Wortfolgestatistik ("Language Modell"), d. h. die im "off-line Dialog Editor" spezifizierten Mehrwortkommandos liefern die erlaubten Wortkombinationen. Im Extremfall beinhaltet die Klassifikation auch die Erkennung und Aussonderung von Füllauten (Äh, Hmm, Räusperer, Pausen) oder "Garbagewörtern" ("Nichtwörtern"). "Garbagewörter" sind sprachliche Ergänzungen, die den eigentlichen Sprachkommandos — unnötigerweise — vom Sprecher hinzugefügt werden, die aber in den Vokabularien des Spracherkenners nicht enthalten sind. Beispielsweise kann der Sprecher das Kommando "preis mit Radius eins" noch erweitern um Begriffe wie "Ich möchte jetzt einen ..." oder "Bitte einen...".

6. Sprecherabhängiger Erkenner

Für die sprecherabhängige Erkennung wird auf derselben Vorverarbeitung aufgesetzt wie für den sprecherunabhängigen Erkenner. Aus der Literatur sind unterschiedliche Lösungsansätze bekannt (z. B. dynamische Zeitnormierung, Neuronale Netz-Klassifaktoren), die ein Training im Echtzeitbetrieb erlauben. Es handelt sich hierbei in erster Linie um Einzelworterkenner, wobei hier vorzugsweise das Verfahren der dynamischen Zeitnormierung zum Einsatz kommt. Um die Benutzerfreundlichkeit zu erhöhen, wird eine Kombination von sprecherabhängigem und sprecherunabhängigem Erkenner im Verbundwortmode verwendet ("Gloria anrufen", "Neues Ziel Onkel Willi", "Funktion Schrägellipse darstellen") wobei die Namen "Gloria", "Onkel Willi", "Schrägellipse" vom Benutzer beim Training frei gewählt wurden und samt den zugehörigen Telefonnummern/Zieladressen/Funktionsbeschreibungen in entsprechenden Listen abgelegt wurden. Der Vorteil dieses Lösungsansatzes liegt darin, daß ein bis zwei (oder noch mehr) Dialogschritte eingespart werden.

7. Nachverarbeitung: Syntax und Semantikprüfung

Das SDS beinhaltet eine leistungsfähige Nachverarbeitung der vom Spracherkenner gelieferten Ergebnisse. Dazu gehören die syntaktischen Prüfungen dahingehend, ob die ermittelten Satzhypothesen dem a priori festgelegten Aufbau der Sprachkommandos ("Syntax") entsprechen. Falls nicht, werden die entsprechenden Hypothesen verworfen.

Weiterhin werden die vom Spracherkenner gelieferten Satzhypothesen auf ihren Sinn und auf ihre Plausibilität überprüft.

Nach dieser Plausibilitätsprüfung wird die aktive Satzhypothese entweder an die Dialogsteuerung weitergereicht oder zurückgewiesen.

Im Falle einer Rückweisung wird die nächstwahrscheinliche Hypothese des Spracherkenners hergenommen und auf gleiche Art und Weise behandelt.

Im Falle eines syntaktisch korrekten und plausiblen Kommandos wird dieses zusammen mit der Beschreibung der Bedeutung an die Dialogsteuerung weitergegeben.

8. Dialog- und Ablaufsteuerung

Die Dialogsteuerung reagiert auf den erkannten Satz und bestimmt die auszuführenden Funktionen. So z. B. legt sie fest

- welche Rückfragen, Informationen oder Aufforderungen an den Benutzer ausgegeben werden,
- welche Aktuatoren wie angesprochen werden,
- welche Systemmodule aktiv sind (sprecherunabhängiger Erkenner, Training)
- welche Teilwortschätze (Teilvokabularien) für die als nächstes erwartete Antwort aktiv sind (z. B. nur Ziffern).

Des weiteren behält die Dialogsteuerung den Überblick über den Status der Applikation, soweit der dem SDS mitgeteilt wird. Der Dialogsteuerung unterlagert ist die Ablaufsteuerung, die die einzelnen Prozesse zeitlich und logisch kontrolliert.

9. Kommunikations- und Kontrollinterface

Hier wird die Kommunikation mit den angeschlossenen Peripheriegeräten abgewickelt.

Dazu stehen verschiedene Schnittstellen zur Verfügung. Das SDS setzt i.a. allerdings nicht alle diese Schnittstellen voraus. Die in der Fig. 2 genannten sind nur Möglichkeiten einer Implementierung.

Das Kommunikations- und Kontrollinterface wickelt insbesondere die Sprachein- und -ausgaben z. B. über

A/D-bzw. D/A-Wandler ab.

10. Spracheingabe/-ausgabe

Die Sprachein/ausgabe setzt sich zusammen aus einem "Sprachsignal-Kompressionsmodul" (= "Sprachencoder"), der die Redundanz bzw. Irrelevanz aus dem digitalisierten Sprachsignal entzieht und somit ein Sprachsignal definierter Dauer in einem erheblich kleineren Speicher als direkt nach der A/D-Wandlung ablegen kann. Die komprimierte Information wird in einem Sprachspeicher abgelegt und für die Ausgabe im "Sprachdecoder" regeneriert, so daß das ursprüngliche eingegebene Wort ohne größeren Qualitätsverlust wieder hörbar ist.

Für die Dialogführung sind im Sprachspeicher bereits von vorneherein ("off-line Dialog Editor") eine Reihe von Kommandos, Hilfstexten oder Anweisungen abgelegt, die den Benutzer bei der Bedienung unterstützen sollen, bzw. ihm Informationen von der Applikationsseite her zukommen lassen.

Weiterhin wird die Sprachcodierung während des Trainings für den sprecherabhängigen Erkenner aktiviert, da der vom Benutzer eingesprochene Namen gleichfalls in den Sprachspeicher kommt. Der Benutzer kann durch Abhören seiner Namensliste jederzeit akustisch über den Inhalt, d. h. die einzelnen Namen informiert werden. Bezüglich Sprachcodier- und -decodieralgorithmus werden Verfahren angewandt, die aus der Literatur bekannt sind und per Software auf einem programmierbaren Prozessor implementiert werden.

In Fig. 6 ist ein Beispiel für einen möglichen Hardwareaufbau der SDS gemäß Fig. 2 dargestellt. Die Konfiguration der einzelnen Funktionsblöcke sowie die Schnittstellen zur Datenein- und Datenausgabe bzw. zur Steuerung der angeschlossenen Komponenten ist hier nur beispielhaft dargestellt. Der hier angenommene aktive Wortschatz (Vokabular) für sprecherunabhängig gesprochene Worte kann z. B. einige hundert Worte umfassen.

Der digitale Signalprozessor (DSP) ist ein handelsüblicher programmierbarer Prozessor, der sich von einem Mikroprozessor durch andere Busarchitektur (z. B. Harvard-Architektur statt Von-Neumann-Architektur), spezielle "on-chip"-Hardware-Rechenwerke (Multiplizierer/Akkumulatoren/Shifter etc.) und I/O-Funktionalitäten auszeichnet, die bei echtzeitiger digitaler Signalverarbeitung benötigt werden. In zunehmendem Maße bieten leistungsfähige RISC-Prozessoren ähnliche Funktionalitäten wie DSPs und können diese ggf. ersetzen.

Der DSP (bzw. ein anderer Mikroprozessor vergleichbarer Leistungsfähigkeit) kann mit Ausnahme spezieller Interface-Kontrollfunktionen sämtliche in Fig. 2 dargestellte Funktionen per Software bzw. integrierter Hardware abwickeln. Mit derzeit handelsüblichen DSPs lassen sich mit dem hier vorgestellten Konzept Wortschätze von z. B. ca. 100 bis 200 Worten realisieren, wobei davon ausgegangen wird, daß dieser Wortschatz vollständig zur Auswahl steht als "aktives Vokabular" und nicht durch Bildung von Teilvokabularien erheblich kleiner ist. Für den Fall, daß Teilvokabularien gebildet werden, kann jedes davon die genannte Größe umfassen.

Durch die Hardwarestruktur gemäß Fig. 6 und insbesondere durch den Verzicht auf zusätzliche Spezialbausteine für die Erkennung und/oder für Dialogabwicklung, Ablaufsteuerung, Sprachkodierung und Interface-Protokollabwicklung bietet sich die Chance einer Realisierung mit einer kompakten, kostengünstigen Hardware mit niedrigem Stromverbrauch. Durch die technologische Weiterentwicklung werden zukünftig höhere Rechenleistungen auf den DSPs verfügbar sein und größere Speicherbereiche adressierbar sein, so daß umfangreichere Vokabularien bzw. leistungsfähigere Algorithmen realisierbar sein werden.

Das SDS wird durch die an den DSP angeschlossene "push-totalk"-Taste (PTT) aktiviert. Ein Bestätigen dieser Taste veranlaßt die Steuersoftware, den Erkennvorgang zu starten. Teile der Signalverarbeitungssoftware sind immer aktiv (Geräuschreduktion, Echokompensation), während die Klassifikation oder die Dialogsteuerung erst durch die PTT aktiviert werden. Im einzelnen sind folgende Module vorhanden:

- A/D- und D/A-Wandler:

Über einen angeschlossenen A/D- und D/A-Wandler werden

— das Mikrofonsignal und ggf. die Lautsprechersignale digitalisiert und zur weiteren Verarbeitung an den DSP übertragen,
— die digitalisierten Sprachdaten zur Sprachausgabe/ Dialogsteuerung in ein Analogsignal zurückgewandelt, verstärkt und an ein geeignetes Wiedergabemedium (z. B. Lautsprecher) weitergereicht.

- D2B optical:

Dies ist ein optisches Bussystem, über welches diverse Audio- und Informationsgeräte gesteuert werden können (z. B.: Autoradio und CD Wechsler, Autotelefon und Navigationsgeräte ...). Dieser Bus überträgt nicht nur Steuer-, sondern auch Audiodaten; im Extremfall (d. h. wenn Mikrofon- und Lautsprechersignal hierüber geschickt werden) erübrigt sich A/D- und D/A-Wandlung im SDS.

- CAN Bus:

Dies ist ein Bussystem, über welches Informationsgeräte und Aktuatoren im Kfz gesteuert werden können; Audioübertragung ist in der Regel nicht möglich.

- V.24-Schnittstelle:

Diese Schnittstelle kann zur Steuerung diverser Peripheriegeräte dienen. Weiterhin kann über diese Schnittstelle die Software des SBS aktualisiert werden. So kann ein entsprechender Wortschatz oder eine entsprechende Sprache (z. B.: Deutsch, Englisch, Französisch ...) geladen werden.

- PCMCIA-Interface:

Diese Schnittstelle dient neben der Kommunikation mit einem Desktop- oder Portable Computer auch der Spannungsversorgung des SDS. Mehrere der oben angeführten Funktionen können hier zusammengefaßt werden. Weiterhin beschreibt diese Schnittstelle neben den elektrischen Eigenschaften auch die mechanischen Abmessungen des SDS. Diese sind z. B. so ausgewählt, daß das SDS in einen PCMCIA-Schacht eines Desktop- oder Portable Computers gesteckt werden kann.

- Speicher:

Der an den DSP angeschlossene Speicher (Daten/Programm-RAM und ROM) dient dem DSP als Programm und Datenspeicher. Ferner beinhaltet dieser die spezifischen Klassifikations-Modelle und ggf. die Referenzmuster für die beiden Spracherkenner und die Festtexte zur Dialogsteuerung und zur Benutzerführung. In einem FLASH- oder batteriegepufferten Speicher werden die benutzerspezifischen Informationen abgelegt (Adress-, Datenlisten).

Funktionsbeschreibung am Beispiel eines sprachbedienten Autotelefons

Im folgenden sind nun die Dialogabläufe exemplarisch am Beispiel einer sprachgesteuerten Telefonsteuerung (z. B. in einem Kfz) beschrieben.

Dieses Beispiel läßt sich erweitern auf die Ansteuerung von Telefon + Radio + CD + Navigation im Kfz bzw. auf die Bedienung eines CAE-Arbeitsplatzes o.a.

Charakteristisch ist für jedes dieser Beispiele:

— Die sprecherunabhängige Erkennung von Mehrwortkommandos, sowie Buchstaben- oder Ziffernkolonnen,
— die sprecherabhängige Eingabe eines vom Benutzer vorher trainierten, freigewählten Namens, dem zugeordnet ist eine Funktion, ein Zahlencode (z. B. Telefonnummer eines Telefonbuches oder Senderfrequenz einer Radiosenderliste) oder eine Buchstabenkombination (z. B. Zielort bei Navigationssystemen).
Bei der Definition der Zuordnung gibt der Benutzer die Funktion, Buchstaben- oder Ziffernkombination im sprecherunabhängigen Verbundwortmode ein (wobei die Funktion, die Buchstaben, Ziffern Teil des zulässigen Vokabulars, d. h. mit "off-line Dialog Editor" vorab festgelegt sein müssen).
— Mit dieser Namenswahl verbunden ist stets die Verwaltung einer entsprechenden Liste unterschiedlicher Namen desselben Benutzers (Telefonbuch, Senderliste, Zielortliste). Diese Liste kann erweitert, gelöscht, abgefragt oder korrigiert werden.

Zustandsdiagramm SDS (Fig. 7)

Während der Bedienung des Telefons über die Spracheingabe nimmt diese unterschiedliche Zustände ein. Die Übergänge werden durch Äußerung von Schlüsselworten gesteuert, wobei die Einleitung einer Äußerung durch die PTT-Taste erfolgt. Ein Gesprächsabbruch erfolgt z. B. durch eine Abbruchtaste.

Betriebszustand "Deaktiviert"

Das Sprachdialogsystem ist in diesem Zustand abgeschaltet.

Betriebszustand "Aktiv" (Fig. 8)

Das Sprachdialogsystem ist aktiviert und wartet nun auf die zur weiteren Steuerung der Peripheriegeräte erlaubten Kommandos. Die Funktionsabläufe des Betriebszustandes "Aktiv" sind in Fig. 8 in Form eines Flußdiagramms (beispielhaft) dargestellt.

Betriebszustand "Namenswahl" (Fig. 9)

Dieser Zustand setzt voraus, daß das entsprechende Schlüsselwort "Namenswahl" bzw. "Telefon Namenswahl" o. ä. richtig erkannt wurde. In diesem Zustand ist die Wahl einer Telefonnummer durch eine Eingabe eines Namens möglich. Dazu wird auf einen sprecherabhängigen Spracherkenner umgeschaltet.

Das Sprachdialogsystem fordert zur Eingabe eines Namens auf. Dieser wird dem Benutzer bestätigt. Das Sprachdialogsystem schaltet nun wieder auf den sprecherunabhängigen Erkenner um.

Sollte der Namen falsch verstanden worden sein, so kann durch eine Abbruchfunktion (Abbruchtaste) ein Wählen der Telefonnummer verhindert werden. Alternativ hierzu ist auch eine Rückfrage des SDS denkbar; ob die dem Sprachkommando zugeordnete Aktion/Funktion ausgeführt werden soll oder nicht.

Der Umfang des Telefonbuches kann je nach Aufwand bzw. Speicherplatz z. B. 50 oder mehr gespeicherte Namen umfassen. Die Funktionsabläufe des Betriebszustandes "Namenswahl" sind in Fig. 9 in Form eines Flußdiagramms dargestellt.

Betriebszustand "Nummernwahl" (Fig. 10)

Dieser Zustand setzt voraus, daß das entsprechende Schlüsselwort richtig erkannt wurde. In diesem Zustand erfolgt die Wahl einer Telefonnummer durch eine Eingabe einer Zahlenfolge. Die Eingabe erfolgt verbunden und sprecherunabhängig.

Der Spracherkenner fordert zur Eingabe einer Nummer auf. Diese wird dem Benutzer bestätigt. Nach der Aufforderung "wählen" wird die Verbindung hergestellt.

Sollte die Nummer falsch verstanden worden sein, so kann durch eine Fehlerfunktion die Nummer korrigiert werden oder über eine Abbruchfunktion, z. B. dem Kommando "Beenden" die Sprachbedienung abgebrochen werden.

Die Funktionsabläufe des Betriebszustandes "Nummernwahl" sind in Fig. 10 in Form eines Flußdiagramms dargestellt.

Betriebszustand "Verbindung"

Die Verbindung ist aufgebaut. In diesem Zustand ist die Spracherkennungseinheit deaktiviert. Das Telefongespräch wird z. B. durch die Abbruchtaste beendet.

Betriebszustand "Nummer speichern/Namen speichern"

Nachdem auf das Schlüsselwort "Nummer speichern" bzw. "Namen speichern" hin das SDS den Benutzer/Sprecher aufgefordert hat, die Ziffern einzugeben und der Benutzer diese eingesprochen hat (vgl. Betriebszustand "Nummernwahl") wird anstelle des Kommandos "wählen" das Kommando "speichern" bzw. ein vergleichbares eingegeben. Die Telefonnummer wird nunmehr gespeichert. Das SDS fordert anschließend den Benutzer auf, den zugehörigen Namen einzusprechen und läßt die Namenseingabe zur Verbesserung des Trainingsergebnisses ein- oder mehrfach wiederholen. Nach dieser Wiederholung ist der Dialog beendet. Zu ergänzen ist, daß die anfängliche Zifferneingabe durch Dialog-Kommandos wie "abbrechen" bzw. "Abbruch", "wiederholen", "korrigieren" bzw. "Korrektur", "Fehler" usw. kontrolliert werden kann.

Betriebszustand "Telefonbuch löschen/Namen löschen"

In Zusammenhang mit dem "Telefonbuch" (Liste aller trainierten Namen und zugehörigen Telefonnummern) sind eine Reihe von Editierfunktionen definiert, die den Komfort des Systems für den Benutzer erhöhen, wie z. B.:

Telefonbuch löschen:

Komplettes oder selektives Löschen, wobei durch Rückfrage ("Sind sie sicher?") des SDS vor dem endgültigen Löschen und durch ggf. Ausgabe des spezifischen Namens ein versehentlichen Löschen durch Erkennungsfehler vermieden wird.

Namen löschen:
Das SDS fordert den Benutzer auf, den zu löschenden Namen zu sprechen. Der Name wird vom SDS wiederholt. Danach wird der Benutzer mit der Frage "Sind sie sicher?" aufgefordert, den Löschvorgang zu bestätigen: Die Eingabe des Sprachkommandos "Ja" veranlaßt das Löschen des Namens aus dem Telefonbuch.

Jedes andere als Sprachkommando eingegebene Wort beendet den Dialog.

Betriebszustand "Telefonbuch anhören":
Das SDS sagt das gesamte Telefonbuch an. Ein Bestätigen der PTT oder die Eingabe eines Abbruchkommandos bricht den Dialog ab.

Betriebszustand "Telefonbuch wählen":
Das SDS sagt das Telefonbuch komplett an. Wird bei dem gewünschten Namen ein Abbruch- oder Wählkommando gegeben bzw. die PTT betätigt, so wird der ausgewählte Namen noch einmal angesagt und nachgefragt "Soll die Nummer gewählt werden?" Die Eingabe des Sprachkommandos "Ja" veranlaßt den Wahlvorgang, d. h. die Verbindung wird hergestellt.

"Nein" veranlaßt das SDS, das Ansagen des Telefonbuches fortzusetzen.

"Abbruch/abbrechen" beendet den Dialog.

Die Eigenschaften des zuvor beschriebenen SDS können wie folgt zusammengefaßt werden:
Benutzt wird ein Verfahren zur automatischen Steuerung und/ oder Bedienung von einem oder mehreren Geräten per Sprachkommandos bzw. per Sprachdialog im Echtzeitbetrieb, bei dem Verfahren zur Sprachausgabe, Sprachsignalvorverarbeitung und Spracherkennung, syntaktisch-grammatikalischen Nachverarbeitung sowie Dialog-, Ablauf- und Schnittstellensteuerung zur Anwendung kommen. Das Verfahren in seiner Grundversion ist im "on-line"-Betrieb durch eine fixierte Syntax- und Kommandostruktur, sowie durch eine Kombination von fixiertem Vokabular (sprecherunabhängiger Erkenner) und frei definierbarem Vokabular, wie z. B. Namen (sprecherabhängiger Erkenner), gekennzeichnet. In vorteilhaften Aus- und Weiterbildungen kann es durch eine Reihe von Merkmalen charakterisiert werden, wonach vorgesehen ist, daß:

- Syntax- und Kommandostruktur während des Echtzeit-Dialogbetriebs fixiert sind,
- Vorverarbeitung, Erkennung und Dialogsteuerung für Betrieb in geräuschbehafteter Umgebung ausgelegt sind,
- für die Erkennung allgemeiner Kommandos, Namen oder Daten kein Training durch den Benutzer erforderlich ist ("Sprecherunabhängigkeit"),
- für die Erkennung spezifischer Namen, Daten oder Kommandos einzelner Benutzer ein Training notwendig ist ("Sprecherabhängigkeit" bei benutzerspezifischen Namen),
- die Eingabe von Kommandos, Namen oder Daten vorzugsweise verbunden erfolgt, wobei die Anzahl der Worte, aus denen ein Kommando für die Spracheingabe gebildet wird, variabel ist, d. h. daß nicht nur Ein- oder Zweiwortkommandos, sondern auch Drei-, Vier- oder Mehrwortkommados definiert werden können,
- eine echtzeitige Verarbeitung und Abwicklung des Sprachdialoges gegeben ist,
- die Sprachein- und -ausgabe nicht nur über Handapparat, Kopfhörer, Headset o. ä., sondern vorzugsweise im Freisprechbetrieb erfolgt,
- die bei Freisprechen im Mikrofon registrierten Lautsprecherechos elektrisch kompensiert werden, um gleichzeitigen Betrieb von Spracheingabe und Lautsprecher (z. B. für Sprachausgabe, Ready-Signale etc.) zu ermöglichen ("Echokompensation"),
- eine laufende automatische Anpassung an die analoge Übertragungscharakteristik (Raumakustik, Mikrofon- und Verstärkercharakteristik, Sprechercharakteristik) im Betrieb erfolgt,
- im "off-line Dialog Editor" die Syntaxstruktur, die Dialogstruktur, das Vokabular und Aussprachevarianten für den Erkenner neu konfiguriert und festgelegt werden können, ohne daß dies zusätzlicher oder neuer Sprachaufnahmen für den unabhängigen Erkenner bedarf,
- im "off-line Dialog Editor" der Sprachumfang für die Sprachausgabe festgelegt wird, wobei
  a) die registrierten Sprachsignale einer digitalen Sprachdatenkompression unterworfen werden ("Sprachcodierung"), anschließend abgespeichert werden und im echtzeitigen Sprachausgabebetrieb nach Auslesen aus dem Speicher eine entsprechende Sprachdecodierung erfolgt, oder b) der Sprachumfang in Form von Text abgespeichert wurde und im echtzeitigen Sprachausgabebetrieb einer "Text to Speech"-Synthese unterworfen wird,
- die Wortstellung durch Vertauschen einzelner Worte eines Kommandos veränderbar ist,
- vorgegebene synonyme Worte nutzbar sind,
- die gleiche Funktion durch Kommandos unterschiedlicher Wortanzahl (z. B. durch Zweiwort- oder Dreiwortkommandos) realisiert werden kann,
- zur Erkennung und anschließender Aussonderung von Einfügungen wie "Äh", "Hm", "Bitte", oder anderer nicht zum Vokabular gehöriger Kommandos dem Nutzvokabular weitere Wörter bzw. Lauteinheiten hinzugefügt werden ("Nichtwörter, Garbagewörter") bzw. Wordspottingansätze genutzt werden,
- die Dialogstruktur durch folgende Eigenschaften sich auszeichnet:
  - flache Hierarchie, d. h. einige wenige Hierarchieebenen, vorzugsweise eine oder zwei Auswahlebenen,
  - Einbindung von "Ellipsen" d. h. Verzicht auf Wiederholung ganzer Kommandosätze mit mehreren Kommandoworten; statt dessen Beschränkung auf kurze Kommandos, z. B. "weiter", "höher", "stärker", wobei dem System aus dem jeweils vorigen Kommando bekannt ist, worauf sich diese Aussage bezieht,
  - Einbeziehung von "Hilfe-" oder "Info-Menüs",
  - Einbeziehung von Rückfragen von seiten des SDS bei unsicheren Entscheidungen des Erkenners ("Wie bitte", "bitte wiederholen", "und weiter"),
  - Einbeziehung von Sprachausgaben, um durch Anregung bestimmter Sprechweisen die Erkennsicherheit zu steigern (z. B. durch die Aufforderung: "bitte lauter"),
- die Spracherkennung durch Betätigung einer "Push-totalk"-Taste ("PTT") aktiviert und dies akustisch quittiert wird (z. B. durch einen "Pieps"-Ton), um anzuzeigen, daß die Eingabe nunmehr erfolgen kann,
- auf die Betätigung der PTT verzichtet werden kann, wenn nach Rückfragen von Seiten der Sprachausgabe im Anschluß daran Spracheingaben

erforderlich sind, wobei die PTT

- entweder Mehrfachfunktionen wahrnimmt oder beinhaltet, z. B. während des Telefonierens ("Auflegen des Hörers", "Abheben des Hörers") bzw. beim Neustart des Sprachdialogsystems bzw. beim Abbruch eines Telefonwahlvorgangs,
- oder ergänzt wird durch zusätzliche Schalter, welche z. B. einen Neustart oder den Abbruch einer Funktion/Aktion erlauben,

— das Dialogsystem eines oder mehrere der folgenden Leistungsmerkmale aufweist:

- die spezifischen (z. B. trainierten) Kommandos, Daten, Namen oder Parameter unterschiedlicher Benutzer werden bei Bedarf für spätere Wiederbenutzung festgehalten,
- vom Sprecher trainierte Kommandos bzw. Namen werden während der Trainingsphase nicht nur der Erkennung zugeführt, sondern auch in ihrem zeitlichen Verlauf aufgenommen, einer Datenkompression ("Sprachkodierung") zugeführt und nichtflüchtig gespeichert,
- die vom Sprecher trainierten Kommandos bzw. Namen werden während der Trainingsphase derart verarbeitet, daß Umgebungsgeräusche während der Aufnahme weitgehend kompensiert werden,

— der Abschluß eines Erkennvorganges optisch bzw. akustisch quittiert wird ("Pieps"-Ton o. ä.) oder alternativ hierzu bei sicherheits- bzw. zeit- oder kostenrelevanten Entscheidungen das Erkennungsergebnis akustisch wiederholt wird (Sprachausgabe) und der Benutzer die Möglichkeit hat, durch ein verbales Kommando oder durch Betätigen eines Schalters (z. B. PTT) die Ausführung der Aktion zu unterbinden,

— das Sprachdialogsystem mit einem optischen Anzeigemedium (LCD Display, Monitor o. ä.) gekoppelt ist, wobei das optische Anzeigenmedium einzelne oder mehrere der folgenden Funktionen übernehmen kann:

- Ausgabe der erkannten Befehle zu Kontrollzwecken,
- Darstellung der vom Zielgerät als Reaktion auf das Sprachkommando eingestellten Funktionen,
- Darstellung verschiedener Funktionen/Alternativen, die per Sprachkommando anschließend eingestellt bzw. ausgewählt oder modifiziert werden,

— jeder Benutzer eigene Namens- oder Abkürzungslisten einrichten kann (vergleichbar einem Telefon- oder Adreßbuch), wobei

- dem vom Benutzer beim sprecherabhängigen Erkenner trainierte Namen eine Ziffernkette, Buchstabenkette oder ein Kommando bzw. eine Kommandosequenz zugeordnet ist, die im sprecherunabhängigen Betriebsmode eingegeben wurde,
- anstelle der erneuten Eingabe der Ziffernkette, Buchstabenkette oder Kommandosequenz der Benutzer die Listenbezeichnung und den von ihm gewählten Namen eingibt, oder neben dem Namen ein geeignetes Kommando eingegeben wird, welches auf die richtige Liste schließen läßt,
- die Liste sprachgesteuert jederzeit um weitere Einträge erweitert werden kann,
- die Liste sprachgesteuert komplett oder selektiv gelöscht werden kann,
- die Liste auf einen Sprachbefehl hin abgehört werden kann, wobei die vom Benutzer eingegebenen Namen und bei Bedarf die zugehörigen Ziffernkette, Buchstabenkette bzw. Kommandos akustisch ausgegeben werden,
- die akustische Ausgabe der Liste zu jedem beliebigen Zeitpunkt abgebrochen werden kann, wobei bei der auf das Kommando "Fehlern, o. ä. bzw. auf das Kommando "wiederholen" folgenden Ausgabe der bisher eingesprochenen Ziffern dieselbe Blockung benutzt wird wie bei der Eingabe,

— eine Folge von Ziffern (Ziffernkolonne) entweder an einem Stück (zusammenhängend) oder blockweise eingesprochen werden kann, wobei

- nach jeder Eingabepause eine Quittierung erfolgt, indem der letzte Eingabeblock von der Sprachausgabe wiederholt wird,
- nach der Quittierung durch ein Kommando "Fehler", "falsch" o. ä. der letzte Eingabeblock gelöscht werden und die verbleibenden, gespeicherten Blöcke akustisch ausgegeben werden,
- nach der Quittierung durch ein Kommando "Löschen" oder eine ähnliche Kommandoeingabe alle eingegebenen Ziffernblöcke gelöscht werden können,
- nach der Quittierung durch ein Kommando "wiederholen" o. ä. die bisher gespeicherten Blöcke akustisch ausgegeben werden können,
- nach der Quittierung durch ein Kommando "Abbruch" oder eine ähnliche Kommandoeingabe die Eingabe der Ziffernkolonne vollständig abgebrochen werden kann,
- nach der Quittierung weitere Ziffern bzw. Ziffernblöcke eingegeben werden können,
- nach der Quittierung die Zifferneingabe durch ein geeignetes Kommando abgeschlossen wird,

— eine Folge von Buchstaben (Buchstabenkolonne) eingesprochen wird, welche zur Auswahl komplexer Funktionen bzw. zur Eingabe einer Vielzahl von Informationen vorgesehen wird, wobei die Buchstabenkolonne zusammenhängend oder blockweise eingegeben wird und

- nach jeder Eingabepause eine Quittierung erfolgt, indem der letzte Eingabeblock von der Sprachausgabe wiederholt wird,
- nach der Quittierung durch ein Kommando "Fehler", "falsch" o. ä. der letzte Eingabeblock gelöscht wird und die verbleibenden, gespeicherten Blöcke akustisch ausgegeben werden,
- nach der Quittierung durch ein Kommando "Löschen" o. ä. alle eingegebenen Buchstaben gelöscht werden können, und im Anschluß daran eine erneute Eingabe erfolgt,
- nach der Quittierung durch ein Kommando "wiederholen" o. ä. die bisher gespeicherten Blöcke akustisch ausgegeben werden können,
- nach der Quittierung weitere Buchstaben bzw. Buchstabenblöcke eingegeben werden können,
- gegebenenfalls ein Abgleich der Buchstabenkolonne mit einer gespeicherten Wortliste erfolgt und daraus das (die) bestpassende(n) Wort (Wörter) extrahiert wird (werden); alter-

nativ hierzu kann dieser Abgleich bereits nach Eingabe der einzelnen Buchstabenblocks erfolgen,

— nach der Quittierung durch ein Kommando "Abbruch" oder eine ähnliche Kommandoeingabe die Eingabe der Buchstabenkolonne vollständig abgebrochen werden kann,

— nach der Quittierung die Buchstabeneingabe durch ein geeignetes Kommando abgeschlossen wird,

— die Ausgabelautstärke der Sprachausgabe und des "Pieps"-Tons den Umgebungsgeräuschen angepaßt sind, wobei die Umgebungsgeräusche während der Sprachpausen bezüglich ihrer Stärke und Charakteristik erfaßt werden,

— der Zugang zum Sprachdialogsystem bzw. der Zugriff auf benutzerspezifische Daten/Kommandos nur durch Eingabe spezieller Kommandoworte bzw. durch Eingabe spezieller Kommandoworte eines autorisierten Sprechers erfolgt, dessen Sprachcharakteristika dem Dialogsystem bekannt sind und von diesem geprüft werden,

— länger andauernde Sprachausgaben (z. B. Info-Menüs) durch gesprochene oder manuelle Abbruchkommandos oder durch die PTT- oder die Abbruchtaste vorzeitig beendet werden können,

— das Sprachdialogsystem in einer der folgenden Formen die manuelle Bedienung obiger Funktionen (z. B. per Schalter, Taste, Drehknopf) ergänzt oder ersetzt:

— die Sprachkommandierung ersetzt keinerlei manuelle Bedienung, sondern existiert neben der manuellen Bedienung, d. h. die Bedienung kann jederzeit manuell erfolgen bzw. weitergeführt werden,

— einige spezielle Leistungsmerkmale sind nur per Spracheingabe aktivierbar, die wesentlichen Geräte- und Bedienfunktionen bleiben sowohl manuell wie per Sprache kommandierbar,

— die Anzahl der manuellen Bedienelemente wird deutlich reduziert, einzelne Tasten bzw. Drehknöpfe übernehmen Mehrfachfunktion. Per Sprache wird manuellen Bedienelementen eine spezielle Funktion zugewiesen. Nur wesentliche Bedienfunktionen sind noch manuell ansteuerbar. Die Basis ist die Sprachkommandierung,

— mit einem einzigen Mehrwortkommando eine Vielzahl unterschiedliche Geräte sowie Gerätefunktionen ansprech- und modifizierbar sind und somit eine umständliche mehrstufige Vorgehensweise (z. B. Auswahl des Gerätes im 1. Schritt, danach Auswahl der Funktion im 2. Schritt, danach Auswahl der Art der Änderung im 3. Schritt) nicht erforderlich ist,

— das Sprachdialogsystem im Kfz für einzelne oder mehrere der im folgenden genannten Funktionen zur Anwendung kommt:

— Bedienung einzelner oder mehrerer Geräte, wie z. B. Autotelefon, Autoradio (ggf. mit Kassetté, CD-Wechsler, Soundsystem), Navigationssystem, Klimaanlage, Heizung, Reiserechner, Beleuchtung, Schiebedach, Fensterheber, Sitzversteller, Sitzheizung, Heckscheibenheizung, Spiegelverstellung und -memory, Sitzverstellung und -memory, Lenkradverstellung und -memory etc.,

— Informationsabfrage von Parametern, wie Öldruck, -temperatur, Wassertemperatur, Verbrauch, Reifendruck etc.,

— Information über notwendige Maßnahmen in besonderen Situationen, z. B. bei hoher Wassertemperatur, geringem Reifendruck etc.,

— Warnung des Fahrers bei Defekten,

wobei

— die sprachgesteuerte Auswahl eines neuen Senders im Autoradio nach einem der folgenden Abläufe erfolgt

— Kommandierung des Suchlaufs auf- bzw. abwärts,

— Spracheingabe der Senderfrequenz vorzugsweise in der umgangssprachlichen Form (z. B. "Einhundertdreikommasieben" bzw. "Hundertdreikommasieben" "Hundertunddreikommasieben" bzw. einschließlich der Frequenzangabe (z. B. "Hundertdreikommasieben MegaHertz"),

— Spracheingabe des gebräuchlichen Sendernamens (z. B. "SDR1"),

— bei der Klimaanlage die gewünschte Temperatur (ggf. nach dem Ort der Fahrgastzelle des Kfz gestaffelt nach links, rechts, vorne, hinten) per Spracheingabe nicht nur relativ, sondern vorzugsweise absolut (d. h. in Grad, Fahrenheit o. ä.) festgelegt werden kann und zusätzlich minimale bzw. maximale bzw. mittlere Temperatur oder die Normaltemperatur kommandiert werden können; ähnlich können die Betriebsbedingungen für das Gebläse im Fahrgastraum festgelegt werden.

— dem Navigationssystem ein Zielort (Ortsname, Straßenname) durch Eingabe von Buchstabenkolonnen im "Buchstabiermode" mitgeteilt wird, wobei auch der Anfang des Namens als Eingabe genügt und das Navigationssystem gegebenenfalls mehrere Kandidaten zur Auswahl anbietet,

— eine oder mehrere der folgenden benutzerspezifischen Namenslisten eingerichtet werden:

— Liste zur Speicherung von Telefonnummern unter vorgebbaren Namen/Abkürzungen,

— Liste zur Speicherung von Zielen für das Navigationssystem unter vorgebbaren Namen/Abkürzungen,

— Liste zur Speicherung von Funktionsnamen für Kommandos oder Kommandofolgen,

— Liste zur Speicherung von Senderfrequenzen des Autoradios unter vorgebbaren Sendernamen bzw. Abkürzungen,

— die Ausgabelautstärke der Sprachausgabe und des "Pieps"-Tons, ggf. auch die Radiolautstärke und die Gebläseeinstellung, unter Berücksichtigung eines oder mehrerer der folgenden Parameter festgelegt werden:

— Fahrzeuggeschwindigkeit,

— Drehzahl,

— Öffnungsbreite der Fenster und des Schiebedaches,

— Fahrzeugtyp,

— Wichtigkeit der Sprachausgabe in der jeweiligen Dialogsituation.

In bezug auf die Vorrichtung zur Realisierung eines Sprachdialogsystems ist u. a. vorgesehen, daß die Ablauf-, Dialog-, Schnittstellensteuerung, die Spracheinausgabe sowie die Sprachsignalvorverarbeitung, Er-

kennung syntaktisch-grammatikalische und semantische Nachverarbeitung mittels Mikro- und Signalprozessoren, Speichern und Schnittstellenbausteinen erfolgt, vorzugsweise aber mit einem einzigen digitalen Signal- oder Mikroprozessor sowie dem erforderlichen externen Daten- und Programmspeicher, den Interfaces sowie den zugehörigen Treiberbausteinen, dem Taktgenerator, der Steuerlogik und den für Sprachein-/-ausgabe erforderlichen Mikrofonen und Lautsprechern samt zugehörigen Wandlern und Verstärkern sowie gegebenenfalls einer Push-to-talk(PTT)-Taste und/oder Abbruchtaste.

Ferner ist vorgesehen, daß über ein Interface

— Daten und/oder Parameter ladbar bzw. nachladbar sind, um z. B. Verfahrensänderungen oder ein Sprachdialogsystem für eine andere Sprache zu realisieren,
— die auf einem separaten Rechner festgelegte oder modifizierte Syntaxstruktur, Dialogstruktur, Ablaufsteuerung, Sprachausgabe etc. auf das Sprachdialogsystem übertragen werden ("off-linie Dialog Editor")
— das Sprachdialogsystem mit mehreren der anzusteuernden Geräte über ein Bussystem und/oder ein ringförmiges Netzwerk verknüpft ist (anstelle von Punkt zu Punkt-Verbindungen zu den einzelnen Geräten) und daß über diesen Bus bzw. das Netzwerk Steuerdaten bzw. Audiosignale bzw. Statusmeldungen des Kfz bzw. der zu bedienenden Geräte übertragen werden,
— die einzelnen anzusteuernden Geräte nicht jeweils ein eigenes Sprachdialogsystem enthalten, sondern von einem einzigen Sprachdialogsystem bedient werden,
— eine oder mehrere Schnittstellen zu Fahrzeugkomponenten oder Fahrzeugrechnern bestehen, worüber permanente oder aktuelle Fahrzeugdaten dem Sprachdialogsystem mitgeteilt werden, wie z. B. Geschwindigkeit,
— das Sprachdialogsystem während der Wartezeit (wo keine Sprachein- oder -ausgabe erfolgt) andere Funktionen z. B. des Radios, des Telefons o.a. übernimmt,
— durch erweiterten Speicher ein multilinguales sprecherunabhängiges Dialogsystem aufgebaut wird, wobei kurzfristig zwischen den Dialogsystemen verschiedener Sprachen umgeschaltet werden kann,
— ein optisches Display mit dem Sprachdialogsystem über ein spezielles Interface bzw. über den Busanschluß gekoppelt ist, wobei dieser Bus vorzugsweise ein optischer Datenbus ist und hierüber sowohl Steuer- wie Audiosignale übertragen werden,
— das vollständige Sprachdialogsystem über eine PCMCIA-Schnittstelle mit der per Sprache zu steuernden Vorrichtung bzw. einem Host- oder Applikationsrechner gekoppelt wird.

Es versteht sich, daß die Erfindung nicht auf die dargestellten Ausführungs- und Anwendungsbeispiele beschränkt ist, sondern vielmehr sinngemäß auf weitere übertragbar ist. So ist es z. B. denkbar, ein solches Sprachdialogsystem zur Bedienung eines elektrischen Wörterbuches oder eines elektronischen Diktier- bzw. Übersetzungssystems zu verwenden.

Eine weitere Ausgestaltung der Erfindung besteht darin, daß

- für relativ begrenzte Anwendungen mit kleiner Syntax die syntaktische Überprüfung in Form eines syntaktischen Bigram-Sprachmodells in den Erkennungsprozeß einbezogen wird und somit die syntaktische Nachverarbeitung entfallen kann,
- bei komplexen Aufgabenstellungen die Schnittstelle zwischen Erkenner und Nachverarbeitung nicht mehr einzelne Sätze, sondern ein sog. "Worthypothesennetz" ist, aus dem in einer Nachverarbeitungsstufe aufgrund syntaktischer Vorgaben mit speziellen Paarungs-Strategien der bestpassende Satz extrahiert wird.

Bezugszeichenliste

SBS Sprachbediensystem
PTT Push-to-Talk
HMM Hidden Markov Modelle
DTW Dynamic Time Warping
CMF Mittelwert befreite Cepstralvektoren
DCT Digitale Cosinus Transformation
FFT Fast Fourier Transformation
LDA Lineare Diskriminanzanalyse
PCM Pulse Code Modulation
VQ Vektorquantisierung
SDS Sprachdialogsystem

Patentansprüche

1. Verfahren zur automatischen Steuerung eines oder mehrerer Geräte durch Sprachkommandos oder per Sprachdialog im Echtzeitbetrieb, bei welchem Verfahren die eingegebenen Sprachkommandos mittels eines sprecherunabhängigen Verbundwort-Spracherkenners und eines sprecherabhängigen Zusatz-Spracherkenners erkannt und gemäß ihrer Erkennungswahrscheinlichkeit klassifiziert werden und dasjenige zu lässige Sprachkommando mit der größten Erkennungswahrscheinlichkeit als das eingegebene Sprachkommando identifiziert und die diesem Sprachkommando zugeordneten Funktionen des oder der Geräte initiiert werden, **gekennzeichnet durch** folgende Merkmale:

- die Sprachkommandos (der Sprachdialog) werden (wird) auf der Basis von mindestens einer Syntaxstruktur, mindestens einem Basiskommandovokabular und bei Bedarf mindestens einem sprecherspezifischen Zusatzkommandovokabular gebildet (geführt);
- die Syntaxstruktur(en) und das (die) Basiskommandovokabular(ien) werden in sprecherunabhängiger Form vorgegeben und sind während des Echtzeitbetriebs fixiert;
- das (die) sprecherspezifische (n) Zusatzkommandovokabular (ien) wird (werden) vom (jeweiligen) Sprecher eingegeben und/oder geändert, indem in Trainingsphasen in- und/oder außerhalb des Echtzeitbetriebs ein nach einem sprecherabhängigen Erkennungsverfahren arbeitender Zusatz-Sprach-erkenner vom (jeweiligen) Sprecher durch ein- oder mehrmalige Eingabe der Zusatzkommandos auf die sprachspezifischen Merkmale des (jeweiligen) Sprechers trainiert wird;
- im Echtzeitbetrieb erfolgt die Abwicklung des Sprachdialogs und/oder die Steuerung des Geräts (der Geräte) wie folgt:
  — vom (jeweiligen) Sprecher eingegebene Sprachkommandos werden einem sprecherun-

abhängigen und auf der Basis von Phonemen arbeitenden Verbundwortspracherkenner und dem sprecherabhängigen Zusatz-Spracherkenner zugeleitet und dort (jeweils) einer Merkmalsextraktion unterzogen und

— im Verbundwortspracherkenner anhand der dort extrahierten Merkmale auf das Vorliegen von Basiskommandos aus dem (jeweiligen) Basiskommandovokabular gemäß der (jeweils) vorgegebenen Syntaxstruktur untersucht und klassifiziert und

— im sprecherabhängigen Zusatz-Spracherkenner anhand der dort extrahierten Merkmale auf das Vorliegen von Zusatzkommandos aus dem (jeweiligen) Zusatzkommandovokabular untersucht und klassifiziert;

— anschließend werden die als mit einer bestimmten Wahrscheinlichkeit erkannt klassifizierten Kommandos und Syntaxstrukturen der beiden Spracherkenner zu hypothetischen Sprachkommandos zusammengefügt und diese gemäß der vorgegebenen Syntaxstruktur auf ihre Zulässigkeit und Erkennungswahrscheinlichkeit untersucht und klassifiziert;

— anschließend werden die zulässigen hypothetischen Sprachkommandos nach vorgegebenen Kriterien auf ihre Plausibilität untersucht und von den als plausibel erkannten hypothetischen Sprachkommandos dasjenige mit der höchsten Erkennungswahrscheinlichkeit ausgewählt und als das vom (jeweiligen) Sprecher eingegebene Sprachkommando identifiziert;

— anschließend wird (werden) die dem identifizierten Sprachkommando zugeordnete(n)

— Funktion(en) des (jeweils) zu steuernden Geräts initiiert und/oder

— Antwort(en) gemäß einer vorgegebenen Sprachdialogstruktur zur Fortführung des Sprachdialogs generiert.

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß die Eingabe von Sprachkommandos manuell und/oder akustisch erfolgt.

3. Verfahren nach Anspruch 2, dadurch gekennzeichnet, daß die Eingabe von Sprachkommandos im Freisprechbetrieb erfolgt.

4. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß akustisch eingegebene Sprachkommandos geräuschreduziert den beiden Spracherkennern zugeleitet werden, indem durch stationäre und/oder quasistationäre Umgebungsgeräusche verursachte Geräuschsignale im Sprachsignal-Empfangskanal vor den beiden Spracherkennern kompensiert werden.

5. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß akustisch eingegebene Sprachkommandos echokompensiert den beiden Spracherkennern zugeleitet werden, indem in den Sprachsignal-Empfangskanal rückgekoppelte Signale einer Sprachausgabeeinheit im Sprachsignal-Empfangskanal vor den beiden Spracherkennern kompensiert werden.

6. Verfahren nach einem der Ansprüche 4 oder 5, dadurch gekennzeichnet, daß die Kompensation mittels adaptiver digitaler Filterverfahren erfolgt.

7. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß die eingegebenen Sprachkommandos nach Digitalisierung blockweise zusammengefaßt und nach einer Gewichtung mittels einer Spektraltransformation in den Frequenzbereich umgesetzt werden und anschließend durch Betragsbildung und nachfolgender gehörbezogener MEL-Filterung zu Kanalvektoren zusammengefaßt werden und daß daran anschließend eine Segmentierung durchgeführt wird.

8. Verfahren nach Anspruch 7, dadurch gekennzeichnet, daß als Spektraltransformation eine Fast-Fourier-Transformation (FFT) eingesetzt wird.

9. Verfahren nach einem der Ansprüche 7 oder 8, dadurch gekennzeichnet, daß die Segmentierung in eine Grob- und eine Feinsegmentierung unterteilt ist.

10. Verfahren nach einem der Ansprüche 7 bis 9, dadurch gekennzeichnet, daß im sprecherunabhängigen Verbundwortspracherkenner die Merkmalsextraktion dergestalt durchgeführt wird,

- daß die Kanalvektoren mit einer diskreten Cosinustransformation (DCT) in Cepstralvektoren transformiert werden,
- daß zusätzlich die Energie des zugehörigen Signals berechnet und normiert wird,
- daß zur Adaption des Erkenners auf den jeweiligen Sprecher und/oder die jeweiligen Übertragungscharakteristiken des Sprachsignal-Empfangskanals fortlaufend der Mittelwert der Cepstralvektoren berechnet und von den Cepstralvektoren abgezogen wird,
- daß die vom Mittelwert der Cepstralvektoren befreite Cepstralvektoren und die berechnete normierte Signalenergie zu mittelwertfreien Cepstralkoeffizienten (CMF-Vektoren) zusammengefaßt werden.

11. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß der sprecherunabhängige Verbundworterkenner bei der Klassifizierung mit einem phonembasierten Hidden-Markov-Modell (HMM) arbeitet.

12. Verfahren nach Anspruch 11, dadurch gekennzeichnet, daß die Klassifikation mit Hilfe eines Viterbialgorithmus durchgeführt wird.

13. Verfahren nach Anspruch 12, dadurch gekennzeichnet, daß der Viterbialgorithmus durch eine vorgegebene Wortfolgestatistik ergänzt wird.

14. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß bei der Klassifikation zusätzlich auch Füllwörter oder -laute oder sonstige im vorgegebenen Basisvokabular nicht enthaltene Fehlkommandos als solche erkannt und entsprechend klassifiziert und ausgesondert werden.

15. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß der sprecherunabhängige Verbundwort-Spracherkenner und der Zusatz-Spracherkenner auf derselben Signalvorverarbeitung für die eingegebenen Sprachkommandos aufsetzen.

16. Verfahren nach Anspruch 15, dadurch gekennzeichnet, daß die Signalvorverarbeitung Verfahren zur Geräuschreduktion, Echokompensation und Segmentierung umfaßt.

17. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß der Zusatzspracherkenner als Einzelwortspracherkenner arbeitet.

18. Verfahren nach Anspruch 17, dadurch gekennzeichnet, daß der Zusatzspracherkenner als Einzel-

wortspracherkenner nach dem Verfahren der dynamischen Zeitnormierung arbeitet.

19. Verfahren nach Anspruch 17, dadurch gekennzeichnet, daß der sprecherunabhängige Verbundwort-Spracherkenner und der sprecherabhängige Einzelwort-Spracherkenner kombiniert im Verbundwortmodus arbeiten.

20. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß während des Echtzeitbetriebs eine fortlaufende Anpassung des Sprachsignal-Empfangskanals an die analoge Übertragungscharakteristik, insbesondere Raumakustik- und/oder Mikrofon- und/oder Verstärker- und/oder Sprechercharakteristik, erfolgt.

21. Verfahren nach einem der Ansprüche 1 bis 20, dadurch gekennzeichnet, daß die vorgegebenen Basiskommandos in sprachcodierter Form vorgegeben und abgespeichert werden und/oder die vom (jeweiligen) Sprecher in Trainingsphasen eingegebenen Zusatzkommandos und/oder im Echtzeitbetrieb eingegebenen Sprachkommandos nach ihrer Eingabe sprachcodiert weiterverarbeitet und/oder abgespeichert werden und daß akustisch auszugebende Sprachkommandos vor ihrer Ausgabe sprachdecodiert werden.

22. Verfahren nach einem der Ansprüche 1 bis 20, dadurch gekennzeichnet, daß die vorgegebenen Basiskommandos und/oder die Zusatzkommandos und/oder die im Echtzeitbetrieb eingegebenen Sprachkommandos in Form von Text abgespeichert werden und daß akustisch auszugebende Sprachkommandos vor ihrer Ausgabe einer Text-zu-Sprache-Synthese unterzogen werden.

23. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß die Syntaxstruktur und die Basiskommandos und/oder die Zusatzkommandos vorab im "offline Dialog Editiermodus" im Labor erstellt und fixiert werden und dem Verbundwort-Spracherkenner in Form von Datenfiles übergeben werden.

24. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß

- die Wortstellung in den Sprachkommandos durch Vertauschen einzelner Worte eines Kommandos veränderbar ist und/oder
- vorgegebene synonyme Worte bei der Bildung von Sprachkommandos nutzbar sind und/oder
- die gleiche Funktion durch Sprachkommandos unterschiedlicher Wortanzahl realisiert werden kann.

25. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß zur Erkennung und anschließender Aussonderung von Einfügungen oder anderer nicht zum Vokabular gehöriger Kommandos dem zulässigen Vokabular weitere Wörter bzw. Lauteinheiten hinzugefügt werden bzw. Wordspottingansätze genutzt werden.

26. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß die Dialogstruktur folgende Eigenschaften aufweist:

- flache Hierarchie mit nur einigen wenigen Hierarchieebenen, vorzugsweise einer oder zweier Hierarchieebenen,
- Einbindung von Ellipsen, bei der Abwicklung des Sprachdialogs,
- Einbeziehung von Hilfe- oder Info-Menüs,
- Einbeziehung von Rückfragen des Sprachdialogsystems bei unsicheren Entscheidungen des Erkenners,
- Einbeziehung von Sprachausgaben, um durch Anregung bestimmter Sprechweisen die Erkennsicherheit zu steigern.

27. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß die Spracherkennung durch Betätigung einer Push-to-talk-Taste (PTT) aktiviert wird oder daß die Spracherkennung durch Betätigung einer Pushto-talk-Taste (PTT) aktiviert und dies akustisch und/oder optisch quittiert wird.

28. Verfahren nach Anspruch 27, dadurch gekennzeichnet, daß der anschließende Sprachdialog bzw. die anschließende Eingabe von Sprachkommandos ohne Betätigung der Push-totalk-Taste abgewikkelt wird.

29. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß das Sprachdialogsystem eines oder mehrere der folgenden Leistungsmerkmale aufweist:

- die spezifischen (z. B. trainierten) Sprachkommandos unterschiedlicher Sprecher werden bei Bedarf für spätere Wiederbenutzung festgehalten,
- vom Sprecher trainierte Sprachkommandos bzw. Namen werden während der Trainingsphase nicht nur der Erkennung zugeführt, sondern auch in ihrem zeitlichen Verlauf aufgenommen, einer Datenkompression ("Sprachkodierung") zugeführt und nichtflüchtig gespeichert,
- die vom Sprecher trainierten Sprachkommandos werden während der Trainingsphase derart verarbeitet, daß Umgebungsgeräusche während der Aufnahme weitestgehend kompensiert werden.

30. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß der Abschluß eines Erkennvorganges akustisch durch einen Kontrollton quittiert wird.

31. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß bei sicherheits- bzw. zeit- oder kostenrelevanten Entscheidungen das Erkennungsergebnis akustisch wiederholt wird (Sprachausgabe) und der Sprecher die Möglichkeit hat, durch ein verbales Kommando oder durch Betätigen der Push-to-talk-Taste die Ausführung der dem Sprachkommando zugeordneten Funktion zu unterbinden oder rückgängig zu machen.

32. Verfahren nach einem der vorhergehenden Ansprüche dadurch gekennzeichnet, daß das Sprachbediensystem mit einem optischen Anzeigemedium (LCD Display, Monitor, Display eines angesteuerten Geräts o. ä.) gekoppelt ist.

33. Verfahren nach Anspruch 32, dadurch gekennzeichnet, daß das optische Anzeigenmedium einzelne oder mehrere der folgenden Funktionen übernimmt:

- Ausgabe der erkannten Sprachkommandos zu Kontrollzwecken,
- Darstellung der vom Zielgerät als Reaktion auf das Sprachkommando eingestellten Funktionen,
- Darstellung verschiedener Funktionen/Alternativen, die per Sprachkommando anschließend eingestellt bzw. ausgewählt oder modifiziert werden.

34. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß jeder Sprecher eigene Namens- oder Abkürzungslisten einrichten kann mit einem oder mehreren der folgenden Merkmale:

- der vom Sprecher beim sprecherabhängigen Erkenner trainierte Namen repräsentiert eine Ziffernkette, Buchstabenkette und/oder ein Kommando bzw. eine Kommandosequenz, die im sprecherunabhängigen Betriebsmode eingegeben wurde,
- anstelle der erneuten Eingabe der Ziffernkette, Buchstabenkette oder Kommandosequenz kann der Benutzer die Listenbezeichnung und den von ihm gewählten Namen eingeben, oder neben dem Namen ein geeignetes Kommando eingeben, welches auf die richtige Liste schließen läßt,
- die Liste kann sprachgesteuert jederzeit um weitere Einträge erweitert werden,
- die Liste kann sprachgesteuert komplett oder selektiv gelöscht werden,
- die Liste kann auf einen Sprachbefehl hin abgehört werden, wobei die vom Benutzer eingegebenen Namen und bei Bedarf die zugehörigen Ziffernkette, Buchstabenkette bzw. Kommandos akustisch ausgegeben werden,
- die akustische Ausgabe der Liste kann zu jedem beliebigen Zeitpunkt abgebrochen werden.

35. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß eine Folge von Ziffern (Ziffernkolonne) entweder an einem Stück (zusammenhängend) oder blockweise eingesprochen werden kann, wobei

- nach jeder Eingabepause eine Quittierung erfolgt, indem der letzte Eingabeblock von der Sprachausgabe wiederholt wird,
- nach der Quittierung durch ein Sprachkommando "Fehler" o. ä. der letzte Eingabeblock gelöscht wird und die verbleibenden, gespeicherten Blöcke akustisch ausgegeben werden,
- nach der Quittierung durch ein Sprachkommando "Löschen" o. ä. alle eingegebenen Ziffernblöcke gelöscht werden können,
- nach der Quittierung durch ein Sprachkommando "wiederholen" o. ä. die bisher gespeicherten Blöcke akustisch ausgegeben werden können,
- nach der Quittierung durch ein Sprachkommando "Abbruch" o. ä. die Eingabe der Ziffernkolonne vollständig abgebrochen werden kann,
- nach der Quittierung weitere Ziffern bzw. Ziffernblöcke eingegeben werden können,
- nach der Quittierung die Zifferneingabe durch ein geeignetes Sprachkommando "Stop" o. ä. abgeschlossen wird,
- durch Eingabe eines eine Aktion/Funktion startenden Sprachkommandos wie "wählen" o. ä. die Eingabe abgeschlossen wird und die dem Sprachkommando zugeordnete Aktion/ Funktion initiiert wird.

36. Verfahren nach Anspruch 35, dadurch gekennzeichnet, daß bei der auf das Sprachkommando "Fehler" o. ä. bzw. auf das Sprachkommando "wiederholen" o. ä. folgenden Ausgabe der bisher eingesprochenen Ziffern dieselbe Blockung benutzt wird wie bei der Eingabe.

37. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß eine Folge von Buchstaben (Buchstabenkolonne) eingesprochen wird, welche zur Auswahl komplexer Funktionen bzw. zur Eingabe einer Vielzahl von Informationen vorgesehen wird, wobei die Buchstabenkolonne zusammenhängend oder blockweise eingegeben wird und

- nach jeder Eingabepause eine Quittierung erfolgt, indem der letzte Eingabeblock von der Sprachausgabe wiederholt wird,
- nach der Quittierung durch ein Sprachkommando "Fehler", o. ä. der letzte Eingabeblock gelöscht wird und die verbleibenden, gespeicherten Blöcke akustisch ausgegeben werden,
- nach der Quittierung durch ein Sprachkommando "Löschen" o. ä. alle eingegebenen Buchstaben gelöscht werden können, und im Anschluß daran eine erneute Eingabe erfolgt,
- nach der Quittierung durch ein Sprachkommando "wiederholen" o. ä. die bisher gespeicherten Blöcke akustisch ausgegeben werden können,
- nach der Quittierung weitere Buchstaben bzw. Buchstabenblöcke eingegeben werden können,
- gegebenenfalls ein Abgleich der Buchstabenkolonne oder der einzelnen Buchstabenblocks mit einer gespeicherten Wortliste erfolgt und daraus das (die) bestpassende(n) Wort (Wörter) extrahiert wird (werden)
- nach der Quittierung durch ein Sprachkommando "Abbruch" o. ä. die Eingabe der Buchstabenkolonne vollständig abgebrochen werden kann,
- nach der Quittierung die Buchstabeneingabe durch ein Sprachkommando "Stop" o. ä. abgeschlossen wird,
- durch Eingabe eines eine Aktion/Funktion startenden Sprachkommandos wie "wählen" o. ä. die Eingabe abgeschlossen wird und die dem Sprachkommando zugeordnete Aktion/ Funktion initiiert wird.

38. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß die Ausgabelautstärke der Sprachausgabe und des Kontrolltons den Umgebungsgeräuschen angepaßt sind, wobei die Umgebungsgeräusche während der Sprachpausen bezüglich ihrer Stärke und Charakteristik erfaßt werden.

39. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß der Zugang zum Sprachdialogsystem bzw. der Zugriff auf benutzerspezifische Daten-Kommandos nur durch Eingabe spezieller Kommandoworte bzw. durch Eingabe spezieller Kommandoworte eines autorisierten Sprechers erfolgt, dessen Sprachcharakteristika dem Sprachdialogsystem bekannt sind und von diesem geprüft werden.

40. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß länger andauernde Sprachausgaben (z. B. Info-Menüs) durch gesprochene oder manuelle Abbruchkommandos vorzeitig beendet werden können.

41. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß das Sprachdialogsystem in einer der folgenden Formen die manuelle Bedienung obiger Funktionen (z. B. per Schalter, Taste, Drehknopf) ergänzt oder ersetzt

- Die Sprachkommandierung existiert neben der manuellen Bedienung, so daß die Bedienung jederzeit manuell erfolgen bzw. weitergeführt werden kann;
- einige spezielle Leistungsmerkmale sind nur per Spracheingabe aktivierbar, die anderen Geräte- und Bedienfunktionen bleiben sowohl manuell wie per Sprache kommandierbar;
- die Anzahl der manuellen Bedienelemente wird deutlich reduziert, einzelne Tasten bzw. Drehknöpfe übernehmen Mehrfachfunktion. Per Sprache

wird manuellen Bedienelementen eine spezielle Funktion zugewiesen. Nur wesentliche Bedienfunktionen sind noch manuell ansteuerbar. Die Basis ist die Sprachkommandierung.

42. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß mit einem einzigen Einzelwort-Mehrwortkommando eine Vielzahl unterschiedliche Geräte sowie Gerätefunktionen ansprech- und modifizierbar sind und somit eine mehrstufige Vorgehensweise nicht oder nur in einem geringen Umfang erforderlich ist.

43. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß das Sprachdialogsystem in Fahrzeugen für einzelne oder mehrere der im folgenden genannten Funktionen zur Anwendung kommt:

- Bedienung einzelner oder mehrerer Geräte, wie z. B. Autotelefon, Autoradio (ggf. mit Kassette, CD-Wechsler, Soundsystem), Navigationssystem, Klimaanlage, Heizung, Reiserechner, Beleuchtung, Schiebedach, Fensterheber Sitzversteller etc.
- Informationsabfrage von Parametern, wie Öldruck, -temperatur, Wassertemperatur, Verbrauch, Reifendruck etc.
- Information über notwendige Maßnahmen, z. B. bei hoher Wassertemperatur, geringem Reifendruck etc.
- Warnung des Fahrers bei Defekten.

44. Verfahren nach Anspruch 39, dadurch gekennzeichnet, daß die sprachgesteuerte Auswahl eines neuen Senders im Autoradio nach einem der folgenden Abläufe erfolgt:

- Kommandierung des Suchlaufs auf- oder abwärts,
- Spracheingabe der Senderfrequenz,
- Spracheingabe des gebräuchlichen Sendernamens.

45. Verfahren nach Anspruch 43, dadurch gekennzeichnet, daß bei der Klimaanlage die gewünschte Temperatur per Spracheingabe relativ oder absolut festgelegt werden kann und zusätzlich eine minimale und/oder maximale und/oder mittlere Temperatur und/oder Normaltemperatur kommandiert werden kann.

46. Verfahren nach Anspruch 43, dadurch gekennzeichnet, daß dem Navigationssystem ein Zielort (Ortsname, Straßenname) durch Eingabe von Buchstabenkolonnen im "Buchstabiermode" mitgeteilt wird, wobei auch der Anfang des Namens als Eingabe genügt und das Navigationssystem gegebenenfalls mehrere Kandidaten zur Auswahl anbietet.

47. Verfahren nach einem der Ansprüche 43 bis 46, dadurch gekennzeichnet, daß eine oder mehrere der folgenden benutzerspezifischen Namenslisten eingerichtet werden:

- Liste zur Speicherung von Telefonnummern unter vorgebbaren Namen/Abkürzungen,
- Liste zur Speicherung von Zielen für das Navigationssystem unter vorgebbaren Namen/Abkürzungen,
- Liste zur Speicherung von Funktionsnamen für Kommandos oder Kommandofolgen,
- Liste zur Speicherung von Senderfrequenzen des Autoradios unter vorgebbaren Sendernamen bzw. Abkürzungen.

48. Verfahren nach einem der Ansprüche 43 bis 47, dadurch gekennzeichnet, daß die Ausgabelautstärke der Sprachausgabe und des Kontrolltons oder der Kontrolltöne, ggf. auch die Radiolautstärke und die Gebläseeinstellung, unter Berücksichtigung eines oder mehrerer der folgenden Parameter festgelegt werden:

- Fahrzeuggeschwindigkeit
- Drehzahl
- Öffnungsbreite der Fenster und des Schiebedaches
- Fahrzeugtyp,
- Wichtigkeit der Sprachausgabe in der jeweiligen Dialogsituation.

49. Verfahren nach Anspruch 28, dadurch gekennzeichnet, daß die Push-to-talk-Taste

- entweder Mehrfachfunktionen wahrnimmt oder beinhaltet, z. B. während des Telefonierens ("Auflegen des Hörers" "Abheben des Hörers") bzw. beim Neustart des Sprachdialogsystems bzw. beim Abbruch eines Telefonwahlvorganges,
- oder ergänzt wird durch zusätzliche Schalter, welche z. B. einen Neustart oder den Abbruch einer Funktion erlauben.

50. Vorrichtung zum Ausführen des Verfahrens nach einem der vorhergehenden Ansprüche, bei welcher eine Spracheingabe-/-ausgabeeinheit über eine Sprachsignalvorverarbeitungseinheit mit einer Spracherkennungseinheit verbunden ist, die wiederum mit einer Ablauf-, Dialog- und Schnittstellensteuerung verbunden ist, dadurch gekennzeichnet daß die Spracherkennungseinheit aus einem sprecherunabhängigen Verbundworterkenner und einem sprecherabhängigen Zusatz-Spracherkenner besteht, die beide ausgangsseitig mit einer Einheit zur syntaktisch-grammatikalischen und/oder semantischen Nachverarbeitung verbunden sind, die mit der Ablauf-, Dialog- und Schnittstellensteuerung verbunden ist.

51. Vorrichtung nach Anspruch 50, dadurch gekennzeichnet, daß die Sprachsignalvorverarbeitungseinheit eine Vorrichtung zur Geräuschreduktion und/oder eine Vorrichtung zur Echokompensation und/oder eine Vorrichtung zur Segmentierung enthält.

52. Vorrichtung nach einem der Ansprüche 50 oder 51, dadurch gekennzeichnet, daß die Spracheingabe-/-ausgabeeinheit einen Sprachencoder, einen Sprachdecoder sowie einen Sprachspeicher enthält.

53. Vorrichtung nach einem der Ansprüche 50 bis 52, dadurch gekennzeichnet, daß die Ablauf-, Dialog- und Schnittstellensteuerung, die Spracheingabe sowie die Sprachsignalvorverarbeitung, Spracherkennung, syntaktischgrammatikalische und semantische Nachverarbeitung mittels mehrerer Mikro- und Signalprozessoren, Speichern und Schnittstellenbausteine erfolgt, oder mittels eines einzigen digitalen Signal- oder Mikroprozessors sowie des erforderlichen externen Daten- und Programmspeichers, der Interfaces sowie der zugehörigen Treiberbausteine, eines Taktgenerators, einer Steuerlogik und der für Sprachein-/-ausgabe erforderlichen Mikrofone und Lautsprecher samt zugehöriger Wandler und Verstärker sowie gegebenenfalls einer Push-to-talk(PTT)- Taste und/oder einer Abbruchtaste.

54. Vorrichtung nach Anspruch 53, dadurch gekennzeichnet, daß über ein Interface

- Daten und/oder Parameter ladbar bzw. nachlad-

bar sind, um z. B. Verfahrensänderungen oder ein Sprachdialogsystem für eine andere Sprache zu realisieren,

- die auf einem separaten Rechner festgelegte oder modifizierte Syntaxstruktur, Dialogstruktur, Ablaufsteuerung, Sprachausgabe etc. auf das Sprachdialogsystem übertragen werden ("off-linie Dialog-Editor").

55. Vorrichtung nach Anspruch 53, dadurch gekennzeichnet, daß diese mit mehreren der anzusteuernden Geräte über ein Bussystem und/oder ein ringförmiges Netzwerk verknüpft ist und daß über diesen Bus bzw. das Netzwerk Steuerdaten und/oder Audiosignale und/oder Statusmeldungen des Kfz und/oder der zu bedienenden Geräte übertragen werden.

56. Vorrichtung nach einem der Ansprüche 50 bis 55 für die Anwendung in Fahrzeugen, dadurch gekennzeichnet, daß die einzelnen anzusteuernden Geräte nicht jeweils ein eigenes Sprachdialogsystem enthalten, sondern von einem einzigen Sprachdialogsystem bedient werden.

57. Vorrichtung nach Anspruch 56, dadurch gekennzeichnet, daß eine oder mehrere Schnittstellen zu Fahrzeugkomponenten oder Fahrzeugrechnern bestehen, worüber permanente oder aktuelle Fahrzeugdaten dem Sprachdialogsystem mitgeteilt werden.

58. Vorrichtung nach einem der Ansprüche 55 oder 56, dadurch gekennzeichnet, daß diese Vorrichtung während der Wartezeiten, in denen keine Sprach-ein- oder -ausgabe erfolgt, andere Funktionen übernimmt.

59. Vorrichtung nach einem der Ansprüche 50 bis 58, dadurch gekennzeichnet, daß durch erweiterten Speicher ein multilinguales sprecherunabhängiges Dialogsystem aufgebaut wird, wobei kurzfristig zwischen den Dialogsystemen verschiedener Sprachen umgeschaltet werden kann.

60. Vorrichtung nach einem der Ansprüche 50 bis 59, dadurch gekennzeichnet, daß ein optisches Display mit dem Sprachdialogsystem über ein spezielles Interface oder über den Busanschluß gekoppelt ist.

61. Vorrichtung nach Anspruch 60, dadurch gekennzeichnet, daß dieser Bus ein optischer Datenbus ist und hierüber sowohl Steuer- wie Audiosignale bzw. Statusmeldungen des Kfz und der zu bedienenden Geräte übertragen werden.

62. Vorrichtung nach einem der Ansprüche 50 bis 61, dadurch gekennzeichnet, daß das vollständige Sprachdialogsystem über eine PCMCIA-Schnittstelle mit dem per Sprache zu steuernden Gerät und/oder einem Host- oder Applikationsrechner gekoppelt wird.

Hierzu 7 Seite(n) Zeichnungen

Nummer: DE 195 33 541 C1
Int. Cl.[6]: G 10 L 7/08
Veröffentlichungstag: 27. März 1997

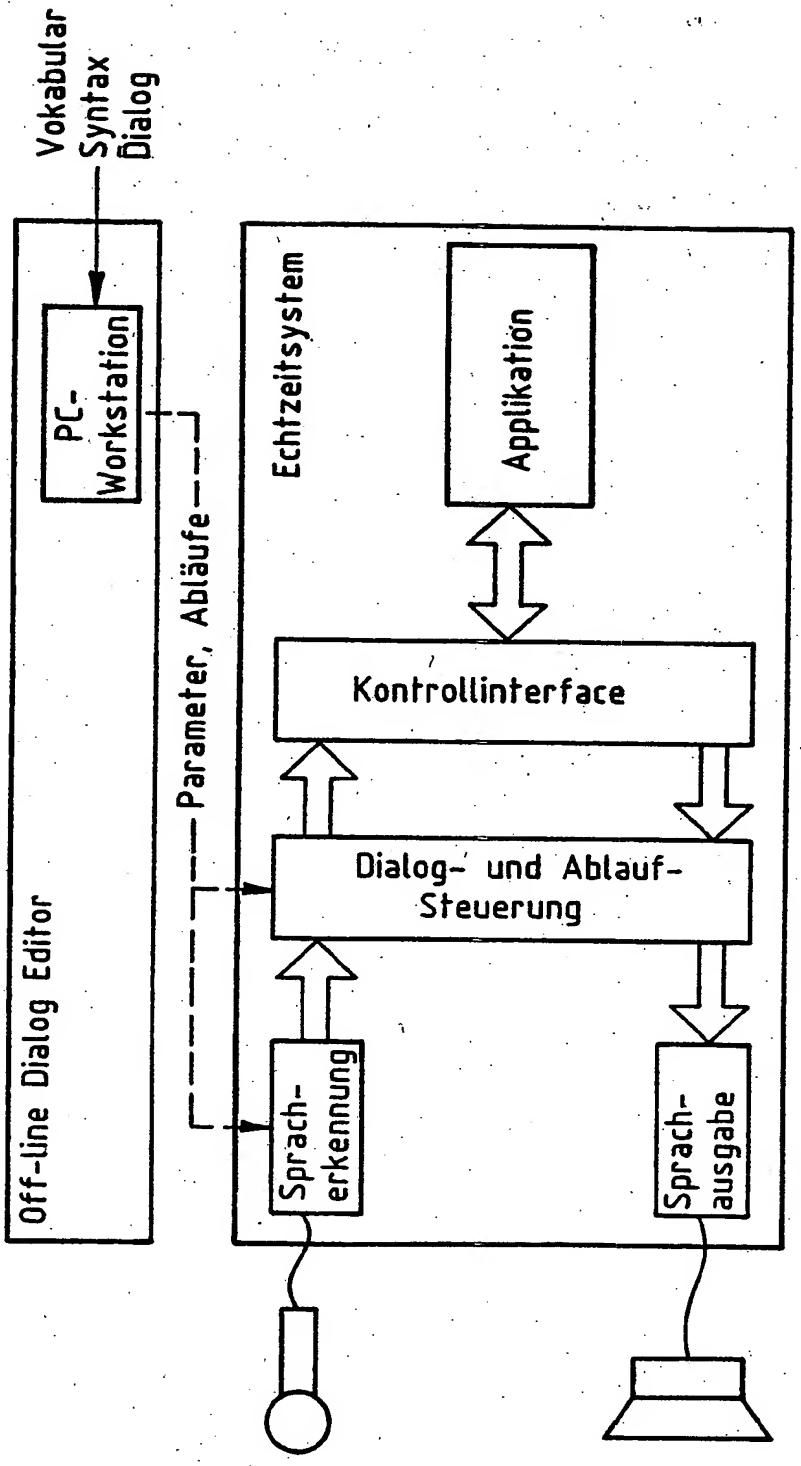

FIG. 1

Nummer: DE 195 33 541 C1
Int. Cl.[6]: G 10 L 7/08
Veröffentlichungstag: 27. März 1997

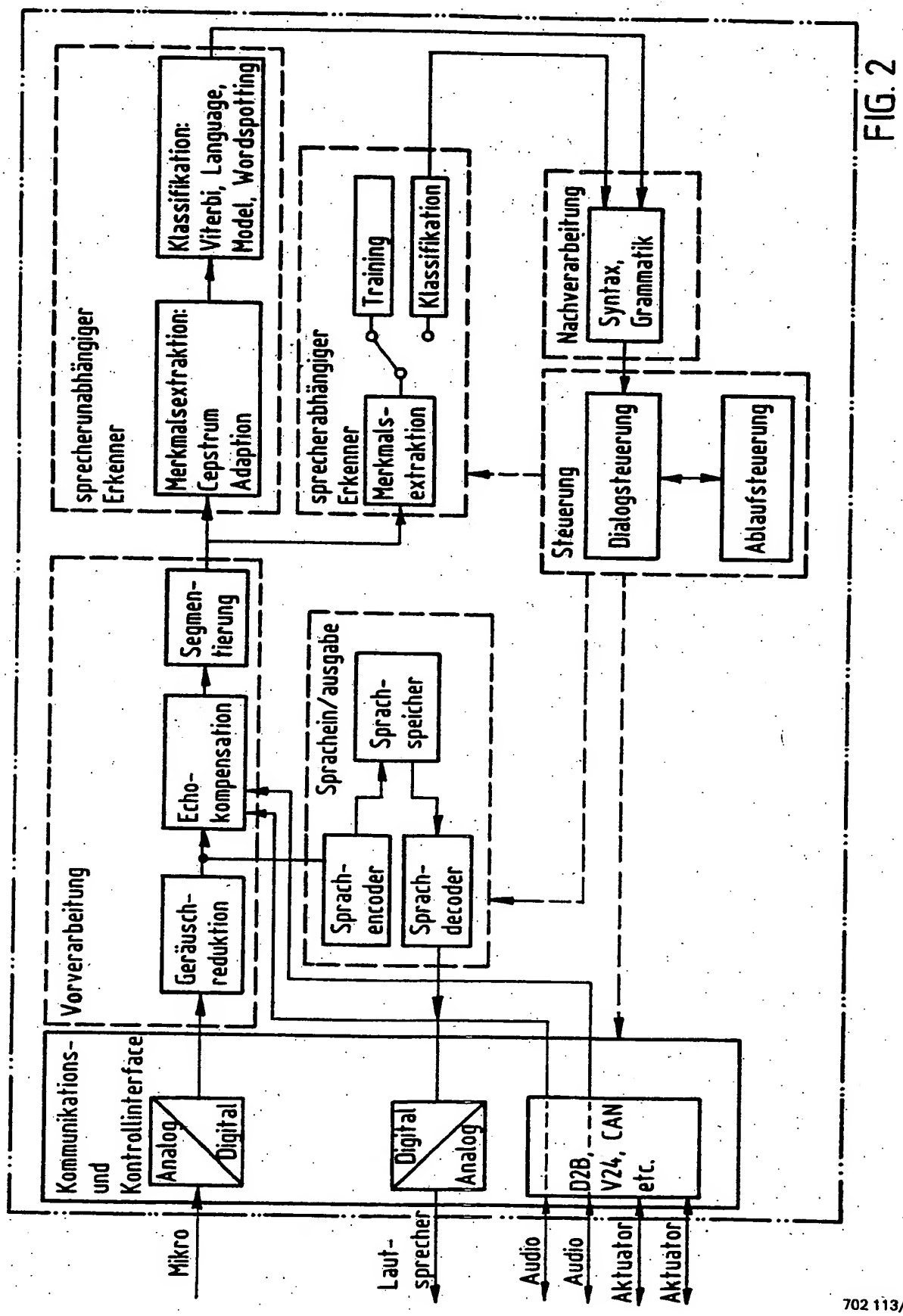

FIG. 2

Nummer: DE 195 33 541 C1
Int. Cl.[6]: G 10 L 7/08
Veröffentlichungstag: 27. März 1997

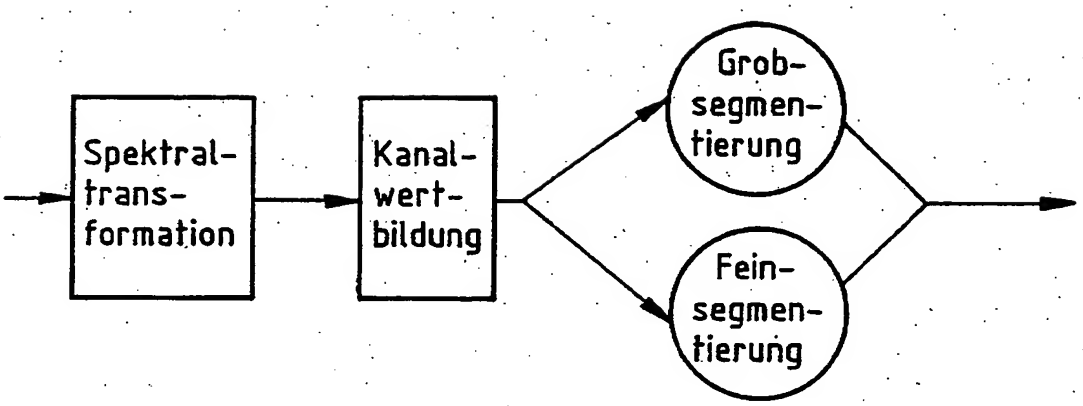

FIG. 3

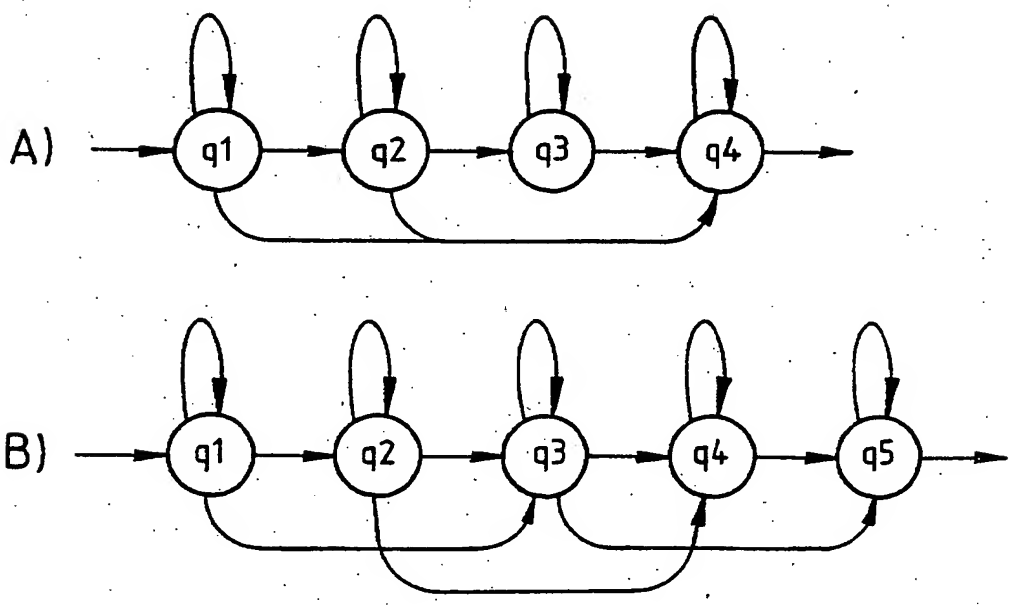

FIG. 4

Nummer: DE 195 33 541 C1
Int. Cl.[6]: G 10 L 7/08
Veröffentlichungstag: 27. März 1997

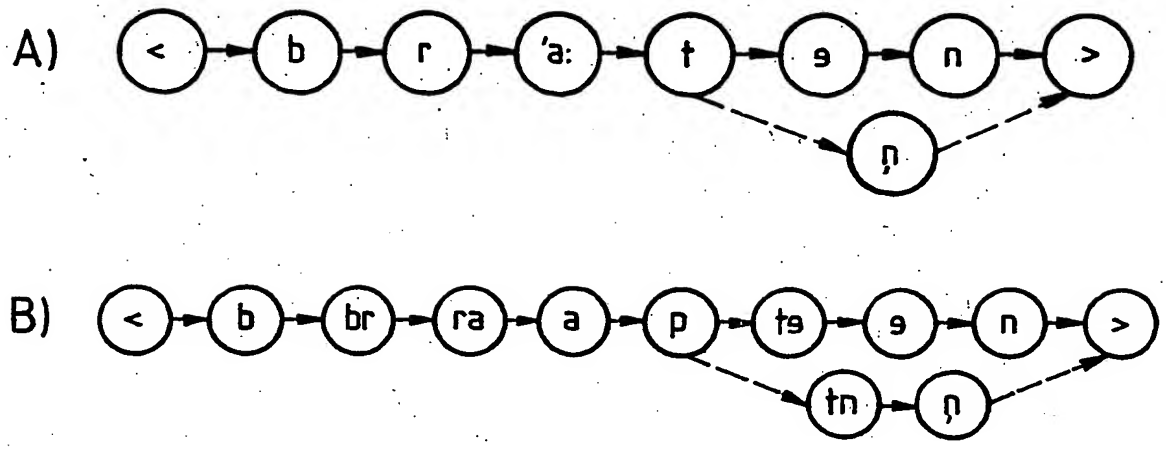

FIG. 5

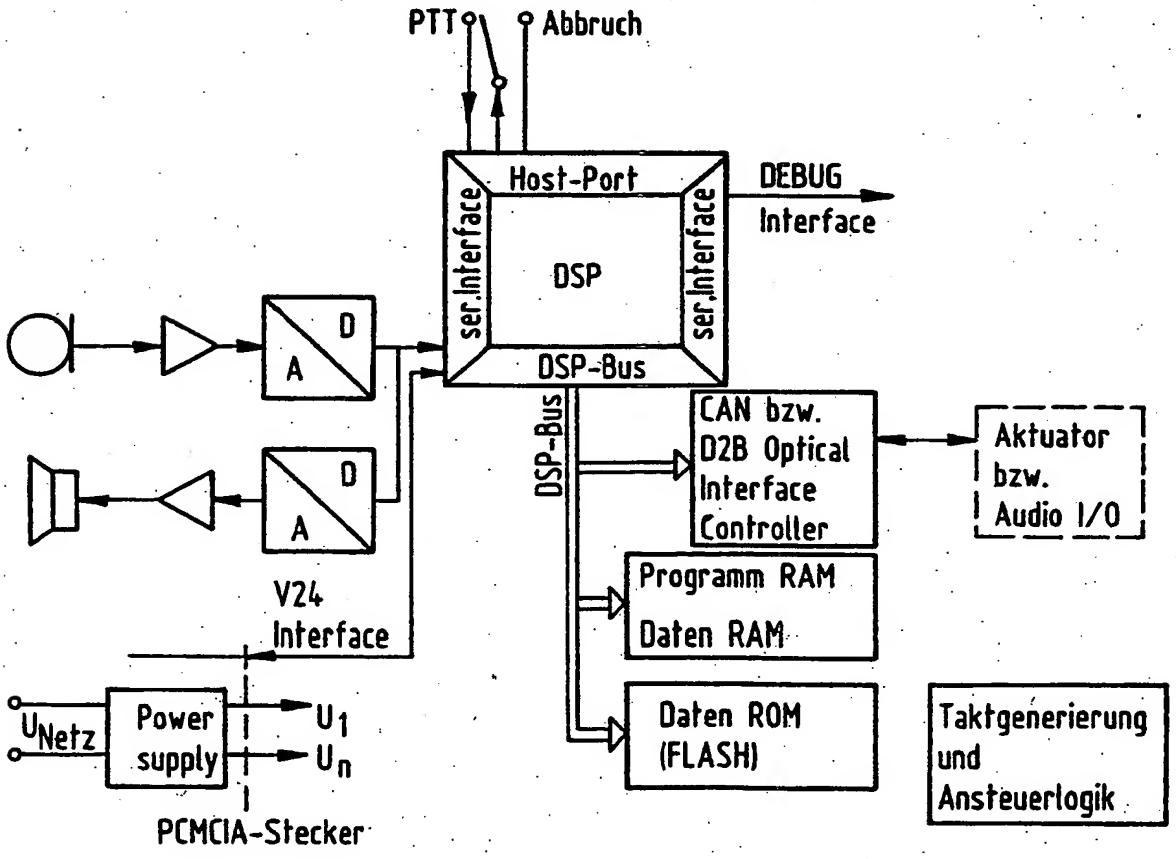

FIG. 6

Nummer: DE 195 33 541 C1
Int. Cl.[6]: G 10 L 7/08
Veröffentlichungstag: 27. März 1997

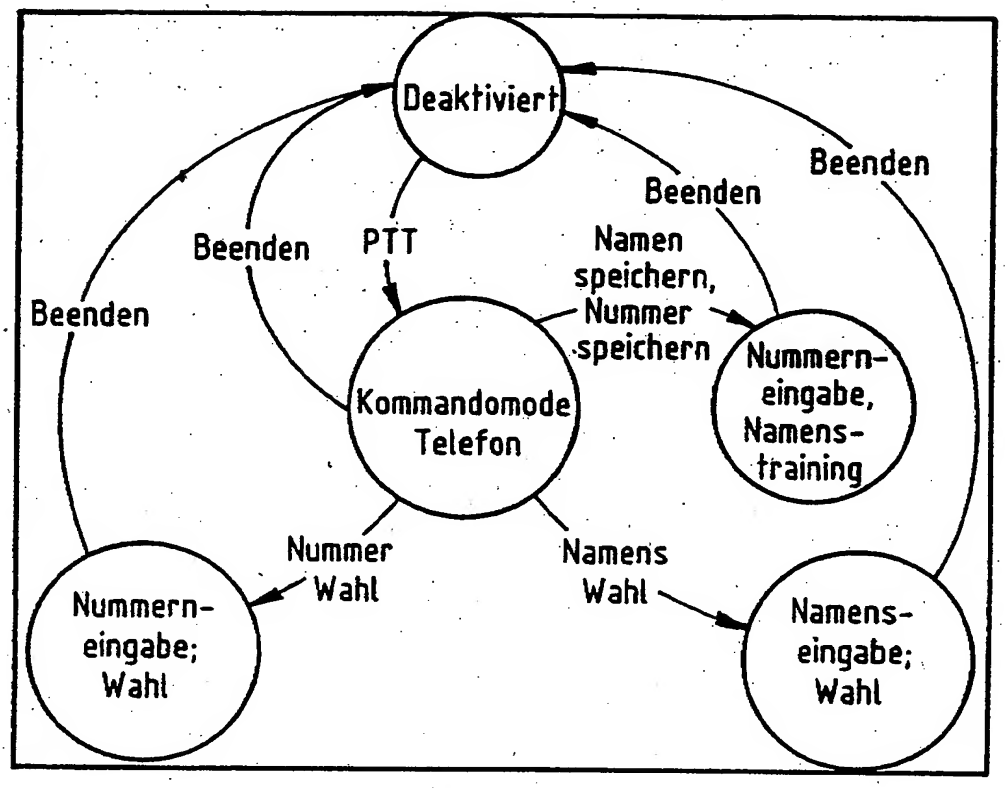

FIG. 7

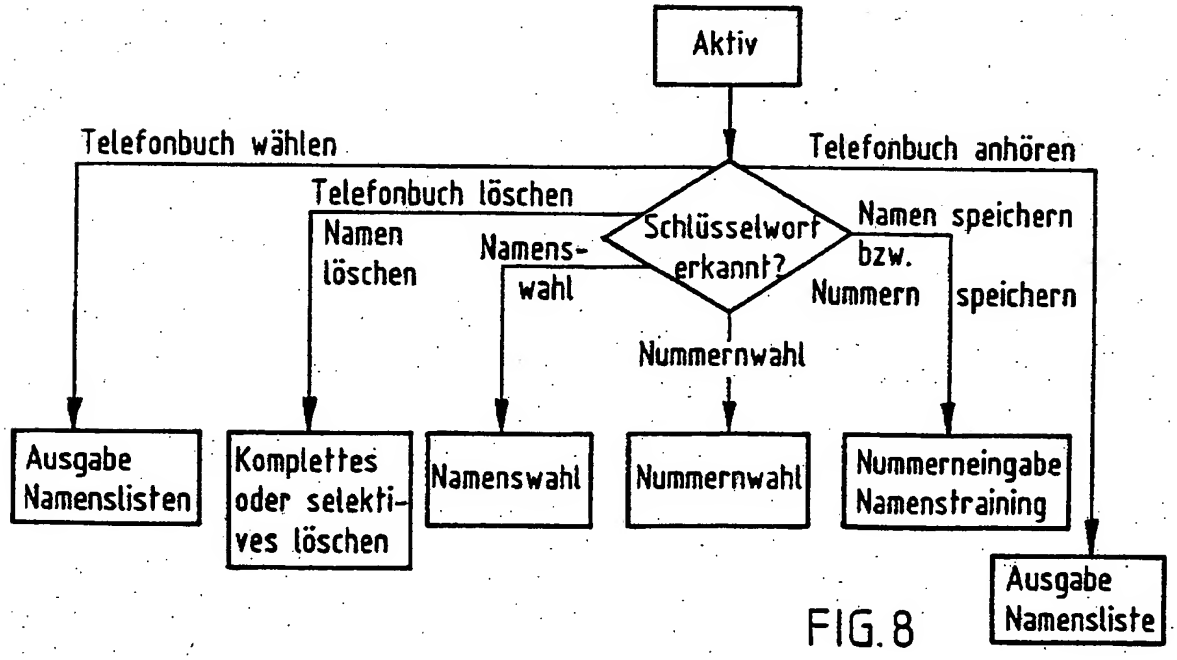

FIG. 8

Nummer: DE 195 33 541 C1
Int. Cl.[6]: G 10 L 7/08
Veröffentlichungstag: 27. März 1997

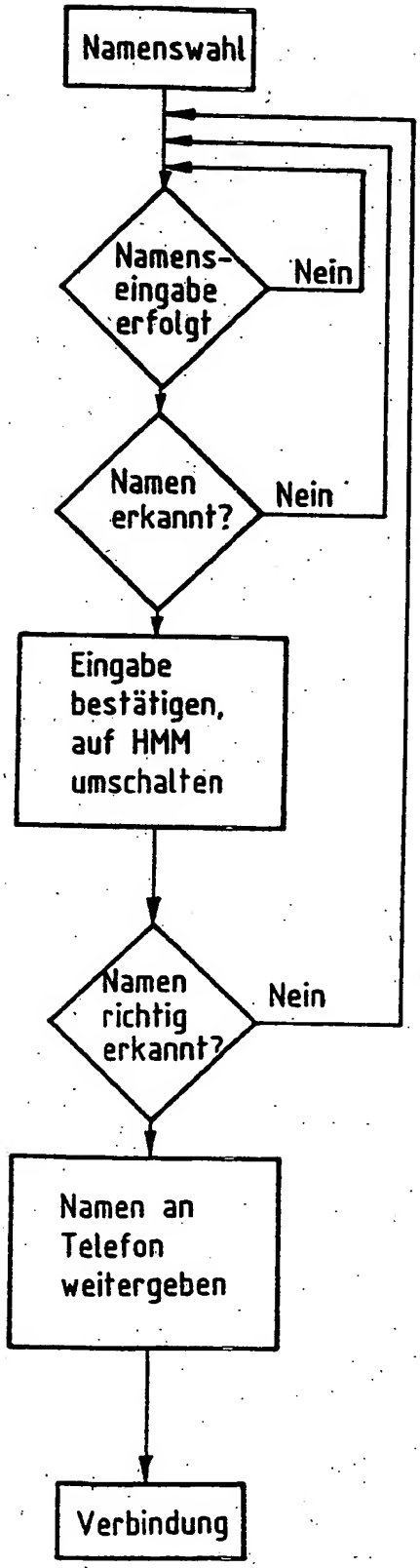

FIG. 9

Nummer: DE 195 33 541 C1
Int. Cl.[6]: G 10 L 7/08
Veröffentlichungstag: 27. März 1997

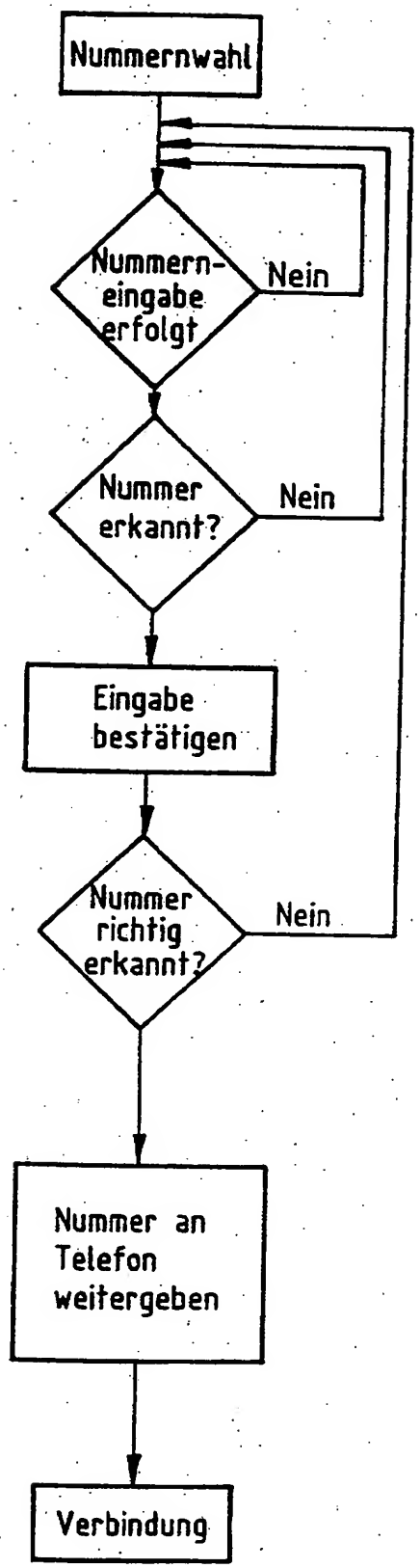

FIG. 10